ihmc
FLORIDA INSTITUTE FOR HUMAN & MACHINE COGNITION

July 27, 2015

ATTN: ONR
675 North Randolph St.
Arlington, VA 22203-2114

Subject: Transmittal Letter for Final Report for N00014-14-1-0086

Title: Accomplishments of Naturalistic Decision Making Research and Applications to Emerging Defense Challenges

IHMC Principal Investigator: Dr. Robert Hoffman

Submitted to: ONR

Dear Dr. Hawkins,

Enclosed is the final report for Accomplishments of Naturalistic Decision Making Research and Applications to Emerging Defense Challenges.

Sincerely,

Diana Thacker

Diana Thacker
Director for Grants and Contracts
IHMC

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 29/07/2015	**2. REPORT TYPE** Final Report	**3. DATES COVERED** *(From - To)* 01/12/2013 – 31/07/2015
4. TITLE AND SUBTITLE Accomplishments of Naturalistic Decision Making Research and Applications to Emerging Defense		**5a. CONTRACT NUMBER**
		5b. GRANT NUMBER N00014-14-1-0086
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Dr. Robert Hoffman		**5d. PROJECT NUMBER**
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Florida Institute for Human and Machine Cognition (IHMC) 40 South Alcaniz St. Pensacola, FL 32502		**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research 875 N. Randolph St. Arlington, VA 22203-1995		**10. SPONSOR/MONITOR'S ACRONYM(S)**
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; distribution is Unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This is a state-of-the-art report on Naturalistic Decision Making (NDM) with respect to military applications. NDM research has focused on the study of decision making in circumstances that involve high stress, and high-risk, uncertainty, and information overload. The report focuses on aligning NDM research with the emerging threats and challenges that confront military operations, especially the challenges expressed by the Human Systems Priority Steering Council. Participants at the 2015 NDM meeting were asked to envision ways in which the NDM paradigm can be applied to address challenges such as emergency response and cyber defense. This Report include a synopsis of the origins, methodology and results from NDM research as well as the papers that were presented at the 2015 International NDM Meeting.

15. SUBJECT TERMS
Naturalistic Decision Making, emerging threats and challenges

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Harold Hawkins
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U		572	**19b. TELEPHONE NUMBER** *(include area code)* 703-696-4323

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

20150804056

Accomplishments of Naturalistic Decision Making Research

and

Applications to Emerging Defense Challenges

Project PI:

Robert Hoffman, Ph.D.
Senior Research Scientist
Institute for Human and Machine Cognition
40 South Alcaniz St.
Pensacola, FLORIDA 32502-6008
(850) 202-4418
rhoffman@ihmc.us

Table of Contents

Executive Summary

This is a state-of-the-art report on Research and theory in "Naturalistic Decision Making" (NDM) with respect to contributions to and prospects for military applications. NDM is a community of practice which has focused on the study of proficient human decision making in circumstances that involve high stress, and high-risk, uncertainty, and information overload. The major accomplishments of the NDM paradigm implied that there would be value in an integrative report aimed at aligning NDM research with the emerging threats, trends and challenges that currently confront military operations. Special emphasis is placed on orienting our nation's resources in the field of cognitive systems engineering to address the challenges expressed by the Human Systems Priority Steering Council. Participants at that meeting were asked to envision ways in which the NDM paradigm can be applied to address current and emerging national, international, and societal challenges:

- Anticipating and adapting to climate change,
- Rapidly and effectively responding to emergencies and natural disasters,
- Countering the spread of radicalism and coping with new forms of regional conflict,
- Rapidly and effectively responding to epidemics,
- Making good decisions in a world of cyber threats,
- Engaging in nation building,
- Protecting utilities, food supply, and infrastructure,
- Helping policy makers and leaders make good decisions on matters of complexity,
- Providing education and health services in distressed nations and regions.

An Appendix in this Report is a synopsis of the origins and core methodology of the NDM paradigm. An Appendix in this Report presents the papers presented at the 2015 International Meeting on Naturalistic Decision Making.

Motivation for the Meeting and This Report

Department of Defense Programs such as the "Combating Terrorism Technical Support Office, the Air Force's "New Optimization and Computational Paradigms for Design under Uncertainty of Complex Engineering Systems" (AFOSR), the Navy's "Future Computing and Information Environment" (Chief of Naval Operations, 2012) and the Army's "Core Competency" programs in decision sciences and human-system integration (ARL) converge on the need for robust methods for information analysis and decision making in circumstances involving problems that are complex and emergent. There is a need for all operational systems and work methods to be adaptive, for coping with unforeseen and dynamic situations. There is a need for automated systems that are resilient.

The 2012 Report to the Assistant Secretary of Defense (Research and Engineering) of the Human Systems Priority Steering Council (Tangney, 2012) lists a number of research focus areas for Joint Forces:

- Improved information Sharing,
- Improved strategic decision making,
- Intelligence Analysis for complex, evolving threats,
- Support for adaptive planning.
- Interactive information displays that adapt to changing needs,
- Models of the decision space that include models of context,
- Automation that supports intuitive interaction,
- Automation that acts as a partner in human-machine analysis teams,
- Automation that creates and maintain representations of users' beliefs, percepts, goals, intentions, and obligations.

The path forward for the creation of such adaptive and resilient automated human-machine work systems is being charted by researchers in the area of Cognitive Systems Engineering, and especially researchers in the area of "Naturalistic Decision Making." These individuals have focused on the study of the knowledge and skills of by experts that allow them to perform with high efficiency when confronted by tough tasks. NDM research has provided invaluable empirical methods and new theories and models.

Background

Origins of the Naturalistic Decision Making Paradigm

In the mid-1980s, Gary Klein and his colleagues conducted a series of studies on professional firefighting (Calderwood, Crandall, and Klein, 1987; Klein, Calderwood, and Clinton-Cirroco, 1986). They developed a cognitive task analysis procedure now called the Critical Decision Method. Since then, the method has been widely used, and with considerable success in revealing the cue patterns that military experts perceive, their reasoning strategies, and the knowledge and skills that distinguish experts from non-experts. The method has been applied in domains as diverse as neonatal intensive care, military command and control, and operations planning and logistics.

These and subsequent findings motivated an evolving paradigm that called itself "Naturalistic Decision Making." This was meant to distinguish a new community of practice from an existing paradigm that had been called "Judgment and Decision-Making" (JDM). Having an origin in the psychology of economic decision making, JDM research tended to involve studies conducted in the academic laboratory using college students as subjects and typically using simplified and rather artificial reasoning tasks. NDM contrasted itself by the study of experts engaging in cognitive work in the "real world."

NDM as a community of practice began with a first conference in 1989 in Dayton Ohio, at which a group of researchers who were studying different professional domains found a common and distinctive set of goals and methods. The shared motivation was to study the decision making by domain experts working at challenging tasks that that are dynamic, ill-structured, and high-stakes (Orasanu and Connolly, 1993). As Gary Klein described it in 1989:

> *The field of JDM has concentrated on showing the limitations of decision makers – that they are not very rational or competent. Books have been written documenting human limitations and suggesting remedies: training methods to help us think clearly, decision support systems to monitor and guide us, and expert systems that enable computers to make the decisions and avoid altogether the fallible humans... Instead of trying to show how people do not measure up to ideal strategies for performing tasks, we have been motivated by curiosity about how people perform well under difficult conditions* (1998, p. 1).

The origins and accomplishments of NDM are presented in more detail in Appendix A.

Goals of NDM Research

NDM researchers have studied reasoning in uncertain and dynamic environments, reasoning in situations where goals come into conflict, reasoning under stress due to time pressure and high risk, and team or group problem-solving (see for example, Flin, Salas, Strub and Martin, 1997; Hoffman, 2007; Schraagen, Militello, Ormerod, and Lipshitz, 2007; Montgomery, Lipshitz, and Brehmer, 2005; Mosier and Fischer, 2010; Salas and Klein, 2001;

Schraagen, 2008; Zsambok and Klein, 1997). This research spans a great variety of domains including piloting, weather forecasting, wildland firefighting, and many others.

A main goal of NDM research is to discover how people actually make decisions in real situations. The goal is not to mold human decision-making into normative or prescriptive models (such as utility analysis or the decision-analytic model) (Cohen, 1993). NDM research has examined the challenges to sensemaking that experts face (e.g., many sources of information, ambiguous and contradictory information, changing conditions); strategies that experts employ for discovering and integrating information; technical support and training for knowledge acquisition.

Methodological Contributions of the NDM Paradigm

A focus of NDM research is on the development and application of cognitive task analysis methods, to reveal the knowledge and skills of experts, support requirements engineering, and inform the design of work systems and interfaces (see Crandall, Klein and Hoffman, 2006). Klein and his colleagues developed a number of methods, in addition to the Critical Decision Method. These too have come to be widely used—the Knowledge Audit, the Cognitive Walkthrough, the Pre-mortem technique, and the Decision-Centered Design approach. These contributions alone make NDM stand out for a significant and far-reaching contribution to the human, military, and technical sciences.

Theoretical Contributions of NDM

NDM has advanced a number of useful ideas about cognition and reasoning, which have impacted cognitive psychology generally as well as applied psychology, ergonomics, and military psychology. The Recognition Primed Decision Making model has been widely applied and has been instantiated computationally. The Data/Frame Theory of sensemaking is gaining traction as a robust model that can capture adaptive and resilient reasoning, well beyond the forms of reasoning for fixed tasks (i.e., normative stage-theoretic models). The Flexecution Model of Replanning offers a depiction of what actually happens in replanning that is more faithful to the empirical complexities than other models of planning (see Klein, 2007).

As NDM matured, advances in theory were made as the empirical understanding grew. Recognizing that NDM, and its paradigm, are by no means focused exclusively on decision making, researchers have studied a variety of high-level cognitive processes such as mental model formation, mental projection to the future, re-planning, coordinating, and maintaining common ground. In recognition of this, a new distinction has been drawn between "microcognition" and "macrocognition" (Klein, Moon and Hoffman, 2006; Klein et al., 2003). These are complementary, with the former focusing on issues of concern in the traditional psychology laboratory (e.g., short-term memory decay, shifts of attention), and the latter focusing on cognition in real world contexts. The microcognitive approach is most appropriate for probing cognition at the millisecond level of causation, rather than in the larger context of on-the-job performance. The designation of the paradigm as Macrocognition is based on an appreciation of two key ideas:

1) Decision-making in real-world contexts is not a single process, but comes in a variety of forms involving differing strategies and differing sequences of mental operations,

2) The effects of context and the important role of sensemaking in problem-solving in real-world situations mean that for the analysis of expert decision making in any given domain one will likely need multiple models, and multiple kinds of models.

A detailed discussion of the history and accomplishments of NDM is presented in Appendix A.

Synopsis of the History of the NDM Conferences

The International Conferences on Naturalistic Decision Making have been held every other year, alternating between North America and Europe. The meetings have been consistently supported by the U.S. Air Force (Air Force Research Laboratory), U.S. Army (Army Research Laboratory, Army Research Institute), and the U.S. Navy (Office of Naval Research). Additional support has come from NASA (Ames Research Center), The European Office of Aerospace Research, the Dutch Ministry of Defense, TNO Defence, Security and Safety, Verzonden van mijn Android-telefoon via Symantec TouchDown, The United Kingdom Ministry of Defence, the Netherlands Ministry of Defence, The *Association pour le Recherche en Psychologie Ergonomique et Ergonomie*, the Human Factors and Ergonomics Society, The University of Aberdeen, the University of Central Florida, San Francisco State University, the University of West Florida, Middlesex University, Aix-Marseille University-the Provence-Alpes Côte d'Azur Region, and from a number of private sector partners, including Aptima, Inc, Charles River Analytics, Cognitive Performance Group LLC, Chi Systems, Macrocognition LLC, MITRE, The Regional Centre for Human Factors-PEGASE Industrial Consortium, ADIMI, the British Computer Society, and the Institute for Human and Machine Cognition.

- The First NDM meeting, held in Dayton Ohio in 1989, was relatively small, consisting of researchers who had discovered a shared interest in "real world" decision making on the part of professionals in diverse domains. The primary product was the edited volume *Decision Making in Action: Models and Methods* (G. Klein, J. Orasanu, R. Calderwood, and C. Zsambok, Editors, 1993).

- The Second NDM Conference was also held in Dayton Ohio, in 1994. The primary product was the edited volume, *Naturalistic decision making.* (C. Zsambok and G. Klein, Editors, 1997).

- The Third NDM Conference was held in Aberdeen Scotland in 1996. The primary product was the edited volume, *Decision making under stress: Emerging themes and applications* (R. Flin, E. Salas, M. Strub, and L. Martin, Editors, 1997).

- The Fourth NDM Conference was held in Warrington Virginia in 1998. The primary product was the edited volume, *Linking expertise and naturalistic decision making.* (E. Salas and G. Klein, Editors, 2001).

- The Fifth NDM Conference was held in Stockholm Sweden in 2000. The primary product was the edited volume, *How professionals make decisions.* (H. Montgomery, R. Lipshitz, and B. Brehmer, Editors, 2005).

- The Sixth NDM Conference was held in Pensacola Florida in May 2003. The primary product was an edited volume titled *Expertise Out of Context* (R. Hoffman Editor, 2007).

- The Seventh NDM Conference was held in Amsterdam The Netherlands in 2005. The primary product was an edited volume titled *Naturalistic decision making and macrocognition.* (J.M. Schraagen, L.G. Militello, T. Ormerod and R. Lipshitz, Editors, 2007).

- The Eighth NDM Conference was held in Pacific Grove California in 2007. The primary product was an edited volume titled *Informed by Knowledge: Expert Performance in Complex Situations* (K.L. Mosier and U. M. Fischer, Editors, 2010).

- The Ninth NDM Conference was held in London England in 2009. The primary produce was the CD, "Proceedings of the 9th Bi-annual international Conference on Naturalistic Decision Making" (W. Wong and N. Stanton, Editors, 2009).

- The Tenth NDM Conference was held in Orlando Florida in 2011. The primary product was the CD, "Proceedings of the 10th International Conference on Naturalistic Decision Making (NDM 2011)" (S. Fiore, Editor).

- The Eleventh NDM Conference was held in Marseille France in 2013. The primary product was the CD "Proceedings of the 11th International Conference on Naturalistic Decision Making (NDM 2013)" (H. Chaudet, L. Pellegrin and N. Bonnardel, Editors).

- The twelfth NDM Conference, titled "NDM 2015," was held in McLean, VA. this report is the first primary produce from this most recent NDM meeting.

The NDM meetings have highlighted presentations for leading scientists and researchers, including Nobel Award winners. The meetings have always highlighted presentations by individuals who bring in fresh perspectives that present views that are complementary to, and sometimes opposed to the NDM stance. These presentations by "welcomed outsiders" have helped to continually reinvigorate the NDM movement and extend its horizons.

The primary products from these meetings have presented research and theory resulting from studies in diverse professional domains, such as fire fighting, health care, offshore oil production, manned space systems, aviation, business management and leadership, human resources and personnel management, criminal justice, professional sports. Topics have included much more than decision making, about a broad spectrum of cognitive work activities including teamwork, cross-cultural understanding, knowledge management, training, and cognitive task analysis methodology. A particular focus in all the conferences has been military psychology and military affairs. NDM research has examined decision making, sensemaking, planning, leadership and other diverse challenges in military activity in areas spanning logistics, command

& control, air operations, UXV control, tactical visualization and decision-aiding, cyberwork, tactical engagement, coalition and joint operations, and many other topics as well.

About the 2015 Meeting

NDM 2015 was held on 9-12 June 2015, hosted by and held at MITRE Corporation in McLean VA. It was attended by 87 individuals, including 44 paper presenters, 11 poster presenters, and the Keynote and Invited Speakers, listed in Table 1.

Table 1. Keynote and Invited Speakers at NDM 2015

SPEAKER	AFFILIATION	TOPIC
Scott Tousley	Deputy Director of the Cyber Security Division, U.S. Department Homeland Security, Office of Science & Technology	Challenges for Decision Making in Cyberdefense
Dr. Alvin Roth	Nobel Award Winner, Craig and Susan McCaw Professor of Economics at Stanford University and the Gund Professor of Economics and Business Administration Emeritus at Harvard University	Market Design as a Process of Adjustments
Marvin Cohen, Ph.D.	Principal Investigator, Perceptronics Solutions	Rethinking NDM
Mr. John Willison,	Director, Command, Power, & Integration, U.S. Army RDECOM CERDEC	Information and Knowledge Management in Systems of Systems Engineering for cyber Defense
CAPT Joseph Cohn	Deputy Director, Human Performance Training and BioSystems Directorate	Representing and Enhancing Intuitive Decision Making: From Individuals to Societies: Progress, Challenges, and Opportunities for Representing Behavior
Tom Ormerod, Ph.D.	Head of Psychology at the University of Sussex, UK	Emerging challenges for NDM: The Case of Security Screening
Chris Baber, Ph.D.	Chair of Pervasive and Ubiquitous Computing, School of Electronic, Electrical and Systems Engineering, University of Birmingham, UK	Emerging Challenges and the "Un-ness" of Events
Simon Henderson	Centre for Cyber Security & Information Systems, Cranfield University, UK	Decision Making for Challenges in Intelligence Analysis and Cyberwork
Michelle Holko, Ph.D.	Booz-Allen-Hamilton Presented on behalf of COL Matthew	Dynamic Threats of Emerging Diseases

	Hepburn, DARPA Biological Technologies Office	
Daphne Ladue, Ph.D.	Research Scientist, Center for Analysis and Prediction of Storms, University of Oklahoma	Decision Making and Climate Change
Jeff Bradshaw, Ph.D.	Senior Research Scientist, Institute for Human and Machine Cognition	Cyber-Physical Threats in the Food Industry: Toward Real-Time Anticipation, Detection, Response, and Recovery
David Woods, Ph.D.	Integrated Systems Engineering at the Ohio State University	Releasing the Adaptive Power of Human Systems

NDM 2015 highlighted sessions on methodology, intelligence analysis, cyberwork, safety, and cultural understanding. the NDM Schedule is presented in Appendix E.

The "Charge" to the Meeting Participants

The "Charge" is presented in the following textbox.

Orienting the NDM Community to Address Emerging Challenges

Major sponsorship for NDM 2015 is coming from the US Department of Defense. Historically, the DoD has funded many research and development projects that were motivated by the NDM paradigm (see Klein, et al., 1993). These include the development of technologies and work methods for information sharing, strategic decision making, and adaptive planning.

Current NDM research and development topics continue to call for NDM inspiration to develop tools, technologies and work methods that are usable, useful and understandable. There is a growing drive to create adaptive and resilient human-machine work systems (see Tangney, 2012). NDM-inspired research will be crucial in achieving these capabilities.

As a Community of Practice, we must envision ways in which the NDM paradigm can be applied to address current and emerging national, international, and societal challenges:

- Anticipating and adapting to climate change,
- Rapidly and effectively responding to emergencies and natural disasters,
- Countering the spread of radicalism and coping with new forms of regional conflict,
- Rapidly and effectively responding to epidemics,
- Making good decisions in a world of cyber threats,
- Engaging in nation building,
- Protecting utilities, food supply, and infrastructure,
- Helping policy makers and leaders make good decisions on matters of complexity,
- Providing education and health services in distressed nations and regions.

These are needs facing humanity at large, but they also have crucial and immediate military implications. Indeed, they entail new types of missions for armed forces. The challenge of human-system integration has historically been a focus of cognitive systems engineering and NDM. But the Emerging Challenges require that we reach beyond that focus.

We ask all interested NDM participants to compose a short (250-500 word) statement addressing these questions:

> *What would it take to motivate and support the NDM community to address the Emerging Challenges?*
> *What elements of the NDM approach and research paradigm are especially pertinent to the Emerging Challenges?*
> *How can the existing NDM methodologies and theories be extended so that they apply to the Emerging Challenges?*

A number of the participants provided substantive responses to the Challenges. A synopsis of those responses is presented next.

Synopsis of Responses to The Challenges Questions:
"Orienting the NDM Community to Address Emerging Challenges"

> *A general science that is both grounded in theory and focused on the applied must simultaneously try to bring the lab into the world and bring the world into the lab.*

This Section of this Report is a synopsis and integration of presentations and discussions of the "Emerging Challenges" Input for this Report was provided by the Keynote and Invited speakers, the participants in the two "Emerging Challenges" sessions, and a number of other NDM attendees and participants.

A focus for NDM 2015 was on how the Emerging Challenges entail new missions and new ways of working for the Armed Forces and for the government, generally. That being said, the Emerging Challenges relate to new complexities, dilemmas, and conundrums for the entire Free World. They impact business, society, and human welfare and safety. All sectors must respond to the changing nature of new forms of conflict and to global uncertainties of environmental disaster, emergencies, climate change, terrorist attack, and the ever-expanding threats and risks associated with cyber crime and cyberwar. Emerging Challenges necessarily involve multi-agency and multi-Department response, at a scale that exceeds current mechanisms and practice, i.e., from emergency services to local and State and national government to international government and multi-national organizations and agencies. Emerging Challenges involve the pursuit of multiple goals (rather than a single, well defined mission objectives) and the goals could be poorly defined, uncertain, and conflicting.

Emerging Challenges can be characterized by their 'un-ness' (Hewitt, 1983):

Unexpected: Indicating problems in the ways in which events can be predicted, primarily in terms of *when* they will occur but also in terms of *where* they will occur.

Unprecedented: Indicating problems in the knowledge-base relating to such events, and the prior experiences that could be brought to bear in dealing with these events.

Unmanageable: Indicating problems in defining and resourcing response to the events.

The NDM community should be able to make significant contributions to Emerging Challenges, such as rapidly and effectively responding to emergencies and natural disasters, countering the spread of radicalism, rapidly and effectively responding to epidemics, making good decisions in a world of cyber threats, helping policy makers and leaders make good decisions on matters of complexity, helping to cushion the retirement of the baby boomer generation and the loss of expertise that it entails, and so forth. While there was a consensus at NDM 2015 that NDM researchers have been and are continuing to engage more in research that pertains to the Challenges, a number of recommendations have been presented to accelerate the applications of NDM to the emerging challenges.

What would it take to motivate and support the NDM community to address the Emerging Challenges? While the NDM community has a strong tradition on military-related work, it is both possible and necessary to encourage non-military work, e.g., in terms of considering disaster relief and crisis management._Supporting (funding) the NDM community will be problematic given that all resources are stretched. This is not due solely to constraints on budgets, however, but more with the niche in which NDM exists. As a community, NDM spans across several disciplines. While research funders claim to support 'inter-disciplinary' work, their interpretation of 'inter-disciplinary' tends to be a stove-piped project team with 'specialists' in discrete areas, believed to be working together but primarily "doing their own thing" and occasionally coming together to integrate results. This is fundamentally different from an area such as NDM, in which researchers and research teams have to work closely in an inter-disciplinary and interdependent manner in order to address practical, real-world problems. This either means changing the perception of the research funders (so that they have a different view of the terms they are using) or changing the ways in which NDM researchers define their discipline and the ways in which they team with other disciplines.

What elements of the NDM approach and research paradigm are especially pertinent to these Emerging Challenges? NDM is concerned with how decisions are made in "real" (as opposed to laboratory) settings. This means that NDM has a tradition of responding to the messy, chaotic and ambiguous nature of real settings and events. While there are other areas that can lay claim to similar experiences, NDM offers a further benefit in the desire to produce generalizable observations and theories that translate across settings. Thus, rather than solving single problems with single solutions, NDM has the potential to develop broader, cross-problem solutions.

How can the existing NDM methodologies and theories by extended so that they apply to the Emerging Challenges? For emerging challenges such as epidemics, disaster response, and threat to the food supply and to utilities, NDM needs to develop better theories and models to address collaborative response (particularly across very large groups with different working practices, agenda and information needs).

Moving Beyond the Study of Domain Experts

NDM has historically focused on the study of individuals who possess expert level knowledge and skill. The study of such individuals is crucial since it serves as a benchmark for performance. There are important reasons, however, to study the entire proficiency scale:

1). Arguably, the military needs highly trained and skilled individuals (experts) in selected domains (pilots, cyberworkers, etc.). On a broad scale and considering the huge variety of military jobs and tasks, *the military needs journeymen*, that is, individuals who are capable of doing competent work unsupervised.

2). While expertise may be a goal, ideal, or benchmark, one cannot fully understand the endpoint (superior knowledge and skill) without knowing how it *develops*. Both in the community of NDM and in the training community, we currently have insufficient knowledge of exactly what happens in the transition from senior apprentice to junior journeyman, and the transition from senior journeyman to junior expert. Methods of *accelerated learning* have been proposed, and there are cases of successful acceleration (see Hoffman, et al., 2014). However, training must be informed by NDM research, and said *research must look at the cognitive features of individuals who are at proficiency levels other than expert.* Another significant training issue is mentoring (Hoffman and Ward, 2015). *We currently have no empirically-validated and robust method for identifying individuals who might become good mentors, or for developing career tracks for selected individuals who might be come expert mentors as well as domain experts.*

3). Laypersons and individuals of all walks of life make decisions concerning challenges such as epidemic outbreaks and severe weather. This was illustrated at NDM 2015 by the case of the tornado outbreak in Oklahoma in summer 2014. The National Weather Service accurately predicted tornados on a particular day, leading to evacuations with positive result—no loss of life although there was significant property damage. On the next day, similar warnings were issued. The credibility and actionability of the previous day's warnings were on everyone's mind, and so on the second day of the outbreak more people responded. The net effect was that regional highways quickly became parking lots, putting a great many people at risk. What had happened? Upon hearing of the tornado warning, and thinking it highly credible due to the previous day's events, many people decided that the first thing to do was to gather their families together, which meant driving (i.e., from work to home, from home to school to pick up the kids, etc.). This event signalled a few things for the National Weather Service: (a) the need for weather forecasters to integrate their forecasting with the emergency responders and (b) the need for the National Weather service to place renewed emphasis on social-behavioral research aimed at understanding how weather information impacts the reasoning and decision making of laypersons.

NDM should shift from studying experts to studying everyone. At NDM 2015, participants presented a number of interesting examples of "naturalistic" (that is, real world) decision making on the part of lay persons. For example, airport security authorities in the UK determined that more security breeches involved individual carrying British passports than Lebanese passports. The reason was that individuals with malicious intent thought it less likely they would be singled out for interrogation if they carried a British passport. Another problem that was discovered was that in interrogating passengers the security screeners were doing almost all the talking and the interviewee passenger merely has to answer "yes" or "no" to questions. In other words the deception detection strategy being employed was nearly useless.

Another example came from emergency response. During an emergency (e.g., a bombing in a public place) how the police wanted to engage in crowd control and in cordon and search differed from how the fire department wanted to engage in crowd control. Police, public transportation authorities, fire departments, health care responders, etc. all declared different kinds of emergencies at the same places but at slightly different times. The lack of interoperability resulted in significant delays in progress for all the stakeholders. Thus, circumstances in which there are multiple agency/actor teams with differing goals, responsibilities and activities are a potentially fruitful area for application of the NDM paradigm and this is agnostic to whether any or even all of the individuals in the distinct agencies are experts.

4). Whether domain expert or not, decision making is constrained by the pragmatics of circumstance and by motivations. One NDM participant offered this anecdote:

> On a recent flight, I realized I was having a medical problem and was feeling dizziness. Absent any idea of what was wrong, I could not suggest anything to the cabin crew. A common problem is passenger alcohol intoxication but the cabin crew would have known that I'd had no alcohol in-flight. Ebola was in the news and the "scare" was at its peak. While I had been nowhere I could possibly contract Ebola, the cabin crew did not ask. I would have had a fever if Ebola or some other nefarious, infectious ailment, but the cabin crew did not ask about a fever. So I was surprised when I was met at the gate by a wheelchair person who was expecting someone who was intoxicated. Presumably, the crew had advised the authorities that I was ill because of intoxication to ensure they did not have to hold passengers and themselves on the airplane until I was cleared. So, that is a problem, right? How are we ever going to solve a problem like this spread of infectious disease if responsible parties cut corners or make wrong decisions for their own convenience?

The moral here is the importance of non-expert decision making. Whether thought of in terms of so-called "cognitive biases," in terms of contextual influences, self-serving motivations, or other determinants, the "naturalistic" study of decision making of all sorts and in all situations is not only wide open for study using the NDM paradigm (Hoffman and Yates, 2005), but is a focus point for potentially important research that relates directly to many of the Emerging Challenges.

This being said, there is still much of importance that we do not know about expertise. For example, we lack a methodology and theoretical foundation for calculating and measuring the value of expertise, from the standpoint of business or government enterprises. What is the cost of the loss of expert knowledge, and how can that be calculated? Such questions are crucial given the immanent retirement of the "boomer" generation and the consequent "grey tsunami of lost expertise (see Hoffman and Hanes, 2003: Hoffman, et al., 2008). One participant at NDM 2015 commented:

> NDM has helped me guiding different companies (Nutrition, Energy, Construction) to keep knowledge and expertise of older employees within the company. It created an environment with more sharing of information and development of an open structure. Better communication between experts and apprentices has been established. In general performance of employees increased with 15 - 20%.

Just as the Emerging Challenges can be described by their "Un-ness," the opportunities for the study of "real world" decision making can be descripted by their "multi-ness": Multiple partners, multiple options, multiple dilemmas, multiple domains, multiple initiatives, multiple goal conflicts, and multiple responsible actors.

How Can NDM Methodologies be Extended to Apply to the Emerging Challenges?

Almost every aspect of the NDM research paradigm is pertinent to addressing the Challenges. NDM research methods should apply across all stages of research, including: analyzing and making sense of the research questions and goals, conceptualizing solutions and research plans, evaluating solutions, implementing solutions, evaluating impact, learning lessons, and shaping future research. Foundations for such activities cold be established first, including the following activities:

(1). Application of CTA methods for eliciting expertise of experts and stakeholders in the Challenge areas (emergency response, cyber, response to severe weather, etc.) Such expertise could potentially help articulate the nature of the challenges better to the broader NDM community, and also support the development of support approaches. Key uses may include unpacking specific challenge dimensions in a form that can be compared across similar dimensions in other challenge domains, to bring-out the generic and pervasive dimensions of the challenge set.

(2) Setting the stage now for a capacity to capture success stories and best practice in NDM research specifically in the Challenge areas, identifying analogs of the challenge components, and seeing which wheels need not be reinvented.

(3). Establish education and training to empower others to apply the NDM paradigm and methodology to solve problems locally (and thus not be dependent on 'NDM consultants'). This will require the development of education and training in the theory and practice of NDM and CTA, and the means for delivering this to others (including to those at the sharp end who are affected by these challenges) such that they can think about their problems in meaningful ways, and develop their own solutions.

Focused cross-domain and cross-discipline teams might be established to address specific parts of the Challenges by applying NDM methodology. Where specificity may arise is when NDM methods become aligned, coordinated or integrated with specialised technical capabilities tailored to some particular aspect of a Challenge. For example:

Visualization

• Visualization of complex systems, and the impact of human sensemaking and behaviour on physical and social systems. Anticipating and exploring potential futures, branches and sequels. Managing information overload and rapid information evolution for operators using these visualizations.

• HCI design methods that support visualization, data manipulation and exploration, perception of relationships and patterns, mental modeling, and moving such structures to other technical systems for additional analysis.

• Enhanced tools for Concept Mapping, for representing and communicating complex problems and solutions. Enhancement may include the ability to manage uncertainty; managing representations of problems and subsequently overlaying solutions; representing different perspectives of the same problem; enhanced automated layout and weighting analysis, etc.

Training and Learning

• Training for enhanced sensemaking and decision making under conditions of organizational and situational complexity. Accelerating the development of proficiency in less experienced decision makers working with these challenges.

• Enhanced organizational learning for exploiting historical data, by enabling a better understanding of human sensemaking and recognitional strategies in historical decision making data. Such learning may underpin effective anticipation of future scenarios and decision outcomes.

Team Processes

• Renewed investigation of issues in work design for command space architectures, which are a crucial element in all the Challenge areas (intelligence, command and control, crisis management, network administration and monitoring, etc.) to optimize collective sensemaking and action generation.

• Significantly better tools for representing complex environments across distributed teams. Sharing more than just a 'common picture' but the history (and learning) behind it, its meaning, areas of uncertainty or ambiguity, and its implications.

• Process redesign for problem solving, reasoning in teams under conditions of high complexity and inherent uncertainty, and anticipation of decision and behavioral consequences.

• Building enhanced representational systems (including concept mapping and domain ontologies) that support the process of thinking in teams, the communication of meaning, and the development of shared understanding—all specifically with respect to the Challenge areas.

Understanding the Adversary

Adversaries are themselves considered to be domain experts who engage in sensemaking and flexecution activities. A specific emerging challenge is to model and thereby anticipate the formulation of adversary intent in both real world and cyber domains, and anticipate how this is translated into action. This includes modeling and anticipating adversary improvisation, adaptation and creativity in both real world and cyber domains. It includes anticipating and detecting the cues of emergent threat in complex settings. NDM researchers should be engaged in red-teaming, exercises, war gaming, and metacognitive critiquing approaches for enhancing sensemaking, and for evaluating proposed courses of action. NDM research might lead to the development of novel means for influencing, shaping, disrupting and inhibiting adversarial sensemaking, decision making, and action, in both real world and cyber environments.

Cultural Understanding

NDM-ers have begun to study decision making in other cultures and assess the utility of NDM methods and CTA when these are used is studies of decision making on the part of individuals from other cultures (Klein, et al., 2014; Rasmussen, Sieck and Hoffman, in press). More research into the validity and reliability of NDM across different cultures; both as a means of understanding the sensemaking and behavior of others, and in order to anticipate the impact of various interventions on others. NDM studies should begin to include emotion, religion, extremism and violence as key variables, fitting such factors into pattern recognition, data-frame models of sensemaking, and macrocognition in groups. Such factors should also reflect cultural differences and cultural specificity (with regard, for example, to patterns of meaning, values, morality, etc.).

Create Methods for "Rapidized Cognitive Task Analysis" (RCTA)

CTA of various methodological types has a significant track record of success at leading to the development of better work methods and technologies (Flin, Salas, Strub, and Martin, 1997; Hoffman, 2007; Klein, Orasanu, Calderwood, and Zsambok, 1993; Montgomery, Lipshitz, and Brehmer, 2005; Mosier and Fischer, 2010; Salas and Klein, 2001; Schraagen, Chipman and Shalin, 2000; Schraagen, Militello, Ormerod and Lipshitz, 2007; Zsambok and Klein, 1997). One of the major requirements for supporting NDM contributions is the level of skill for undertaking NDM projects. It takes considerable skill to master the CTA observation and interviewing strategies used in NDM research. Unlike the more classical academic decision frameworks, NDM is situated in natural settings rather than in universities. There are few if any college courses on NDM methods such as Cognitive Task Analysis. There are very few opportunities for new NDM researchers to learn the methods and to practice them under supervision. There are no methods for qualifying NDM researchers on their skill at conducting observational and/or interview studies. So skill acquisition is an issue. Quality control is an issue. Cognitive Task Analysis can take considerable time and require considerable resources, not the least important of which is the time required of the domain experts—who in an ideal world would spend all their time doing their jobs and not engaging in as participants in CTA. Ideas about how to "rapidize" CTA have been presented (Zachary, et al., 2012). *What is called for is a research program specifically and explicitly aimed at supporting research to test proposals for rapidizing the process of CTA and validating the rapidized methodologies.*

The Need for Formal (Computational) Models

Empirical research on expertise and cognitive work has conclusively demonstrated that robust decision making depends on "macrocognitive" phenomena at the meaning-level, the knowledge-level, and the context-level. Cognitive work in complex contexts involves certain primary, goal-directed functions including decision-making, sensemaking, re-planning, anticipatory thinking, adapting, detecting problems, and coordinating. Supporting these are high-level cognitive and social processes including maintaining common ground, developing mental models, managing uncertainty, identifying leverage points, and managing attention. Operating in combination, different primary functions and different supporting processes are critical to cognitive work depending on the domain, the particular task, and context. (See the discussion of the microcognition vs. macrocognitive distinction in Appendix A of this Report.)

The concept of macrocognition poses the challenge of how to fit (rather than force-fit) high-level cognition into computational processes and representations, which tend to be reductive. Macrocognition's basis in evidence concerning real-world decision making bolsters the argument that it is necessary for computational models to address cognition at the meaning-level, the knowledge-level, and the context-level, in order for a model to be psychologically plausible. By implication, macrocognitive modeling is necessary in order to allow for the creation of software support tools that are usable, useful, and understandable and that actually help decision makers accomplish their cognitive work. Thus, a prime goal of macrocognition research should be to inform computational cognitive modeling.

It is perhaps not surprising that the first attempt to develop a computational instantiation of an NDM-inspired cognitive model was an attempt to build a model of Recognition Primed Decision Making (see Warwick and Hutton, 2007; see also Fan, McNeese and Yen, 2010). Perhaps it is also not surprising that more attempts have not been made considering that NDM models are not classical, that is, not composed primarily of causal input-output chains. Rather, NDM models emphasize the parallelism and interdependence of macrocognitive processes and functions (see Hoffman, 2010; Hoffman, Klein and Schraaagen, 2007; Klein and Hoffman, 2008; Klein, et al., 2003).

The NDM empirical foundation suggests ways in which human decision making is adaptive and robust, and points out the limits of different adaptive strategies and the individual, team, and organizational barriers to robust decision making. This empirical foundation should be the benchmark capabilities for mathematical and computational modeling. However, in conception the macrocognitive processes are parallel and highly interacting. This feature points directly to limitations in our current ability to computationally model cognitive performance since the models, even those with some parallel processing, rely most heavily on a serial, causal chain approach.

NDM researchers should escape their apparent unease with formal modeling—which discourages attempts to implement NDM ideas in software. Doing so would certainly make NDM more attractive to government sponsors. It might improve NDM theorizing as well. To build formal models, NDM would have to clarify fuzzy macrocognitive concepts, such as framing and story-building, and decision making itself, by making more granular commitments. Application to "real world" problems is central to NDM, but avoidance of formal modeling is not.

A great many schemes and approaches have been devised to conduct computational microcognitive modeling (ACT, GOMS, SOAR, EPIC, many others), and each basic approach has spawned a great many variants and spin-offs. These computational models are often described as being like programming languages in that they allow one to create a model of a task and then run it to produce a step-by-step trace of the cognitive operations that are involved in performing the task. Operations include sensory encoding of stimuli, encoding in memory, executing motor commands, etc.—the so-called "atomic components of thought." Operations have associated with them a time parameter and an error parameter, allowing the model to be used to predict performance times and error rates. More meaningful aspects of cognitive work (e.g., multiple reasoning strategies, knowledge-based reasoning, decision quality) are not captured, modeled or predicted.

The computational microcognitive modeling approach has met with considerable success in:

- Identifying usability problems with new software tools and new interfaces,
- Estimating the cost of training,
- Evaluating alternative designs for interfaces,
- Suggesting ways of improving on software to decrease task execution times, and

- Forming the basic framework for training aids and intelligent tutoring systems that can predict the errors students are likely to make given their stage of their skill development.

However, there are significant challenges to computational microcognitive modeling raised by macrocognition (high-level cognitive processes are parallel and highly interacting) and especially cognitive work in complex systems, where one must consider adaptation, opportunism, dynamics, and the unexpected—rather than routine, well-learned, separable, tasks. When a human who is working on a tough decision problem in context has to deviate from known task sequences to engage in problem solving, collaborative problem solving, or similar activities, then the available models become less applicable. And yet, warfighters at all echelons are confronted more and more with tasks that involve dynamics, uncertainty, and novelty.

It has long been a goal of mathematical psychology and quantitative neuroscience to generate 'grand unified models' of human cognition, and to the present time we have some few examples of rather limited success. These have largely centered on the idea of rational decision-making behaviors and fixed tasks, which naturally lend themselves to mathematical representations. However, we must broach the discovery challenge, the difficult threshold of modeling additional factors that are tied to resilience and adaptation.

As was mentioned above, Recognition-Primed Decision Making (RPD), was the subject of the first attempt to create a computational model (Warwick and Hutton, 2007). The RPD can be taken as an example of a cognitive function that helps make decision making robust. The RPD theory grew out of attempts to understand how decision makers generate and compare multiple courses of action when they are coping in uncertain and dynamic environments (Klein, Calderwood, and Clinton-Cirocco, 1988). Given the burden that such circumstances would impose on the decision maker, it was unclear how experts would be able to make consistently good decisions in real-world environments. The answer hinged on the finding that expert decision makers do not employ analytic decision making strategies (i.e., utility analysis). Instead, experts typically recognize a single course of action based on their experience. Typically, the first recognized course of action would be deemed workable and immediately implemented; but if, for some reason, a shortcoming was detected, another course of action would be generated and considered.

When the relative quality of the different possible courses of action is not obvious, experts begin to use mental simulation to test one option after another to explore the possible results of decisions (Phillips et al., 2004). There are, of course, limits to the variations that can be considered intuitively under emergency time pressures, and even when there are no such pressures (Klein and Brezovic, 1986). Moreover, as the number of viable options become overwhelming, unaided decision makers may simply default to the easiest choice to implement rather than make an otherwise satisfactory choice. For example, a study of more than 800,000 people choosing investment fund options for employee 401(k) plans showed that participation rates fell as the number of fund options increased (Sethi-Iyengar et al., 2004).

Thus, a view of decision making emerged with experience of the decision maker rather than his analytical skill driving the quality of the decision making.

Figure 1 depicts the decision-making process of the RPD computational model at a functional level.

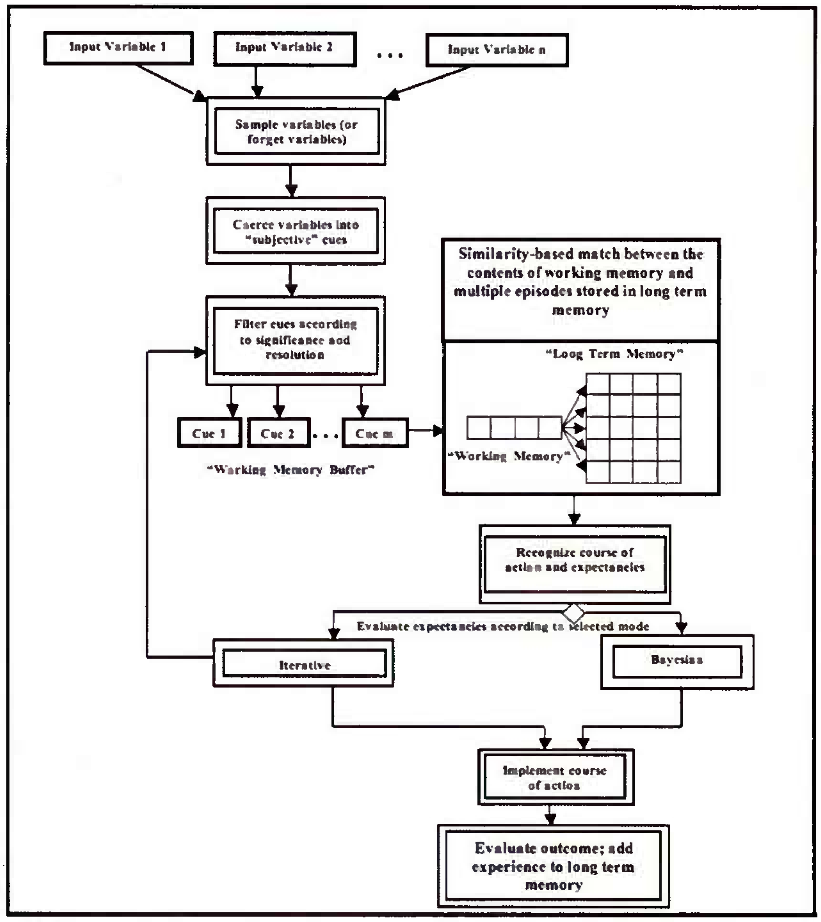


Figure 1. General architecture of the first attempt to model a macrocognitive decision process.

In the most basic terms, the computational model takes variables from an environment as inputs and produces actions as outputs to be implemented in the simulation. But before anything can happen, the computational model must be "populated" with the cues, expectancies, and courses of action that characterize a particular decision in a particular situation. The computational RPD model relies on multiple-trace theory (Hintzman, 1986) and represents a set of episode tokens and types, expressed in terms of the situation that prompted a decision (encoded as a cue vector), the course of action (COA) taken and an outcome measure (either successful or not). A dot product is computed between the vector representing a new situation and each remembered situation in memory. The resulting similarity value is then raised to a user-defined power to determine the proportionate contribution (either positive or negative, according to the outcome of that episode) that each remembered episode makes to a composite recollection. The result is a distribution of recognition strengths across the available course of action, which can then be analyzed in any number of ways to produce output corresponding to a specific course of action. One novel aspect to the model is that COA implementation can itself be represented as an episode, evaluated and stored for use in subsequent decisions. The computational model of RPD has promise for predicting performance at the task of generating courses of action.

Although the RPD computational architecture is generic in the sense that it can represent a variety of decisions, it is specific in the sense that individual instantiations of the model must be created for each type of decision being represented. The model itself makes clear some of the outstanding challenges, such as the generation of COAs.

Considerable progress can and indeed must be made in moving additional NDM conceptual models into the computational realm—such as the Data-Frame model of sensemaking and the Flexecution model of re-planning. These two models are portrayed in Figure 2.

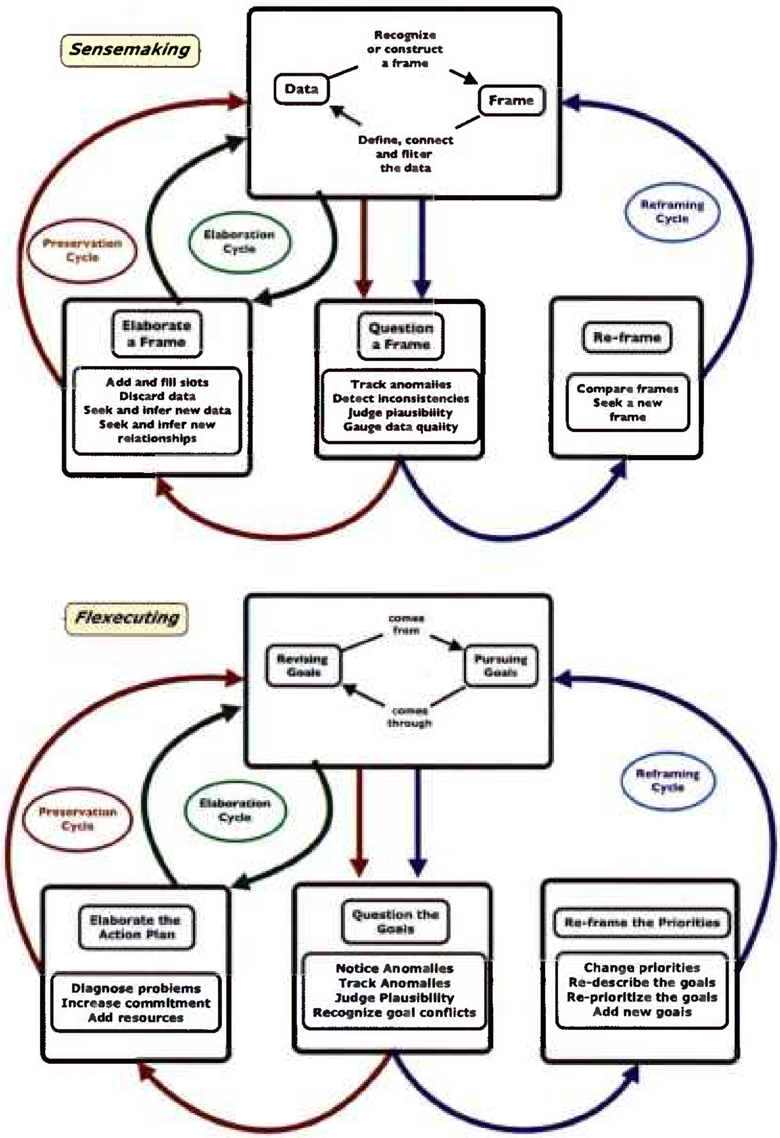


Figure 2. The Data-Frame model of sensemaking and the Flexecution model of Re-planning.

The nodes in the bottom tiers in the Data-Frame and Flexecution Models identify supporting processes that could be implemented as decision aids. There would be significant challenges in implementation a notion of frame (for sensemaking) and in generating an ontology and representation scheme for goals and for tracking goal pursuit (for flexecution).

Such models hold promise as notional architectures for genuine decision aids. That is, they emphasize the patterns that inform decision making and the kinds of issues and considerations that play into decision making (e.g., salient information that indicates anomalies, circumstances calling for the questioning of a frame, the need to evaluate goals while those goals are being pursued, etc. (Whereas most so-called decision aids are actually process control tools in that they control the sequence of activities in which the decision maker must engage.)

A number of other NDM-inspired models can be a focus for attempts at computational modeling.

The Data-Frame Theory of Sensemaking This model emphasizes processes of mental model formation, frame elaboration and re-framing in support of adaptation robust decision-making. Descriptions of a number of possible paths in sensemaking (e.g., questioning a frame followed by re-framing) must be taken further toward computational instantiation (Hoffman and Militello, 2007; Klein, Moon, and Hoffman, 2006a,b).
The "Flexecution" Theory of Re-planning This is a recent outgrowth of the literature in Artificial Intelligence, in which planning has come to be seen as providing support for continuous planning, or re-planning. Flexecution goes further to regard re-planning as a process of in which one discovers goals at the same time as trying to reach them. This leverages the true functions of plans—plans as tools to help one perceive when to be surprised. Available descriptions of possible paths to understanding (e.g., questioning a frame followed by re-framing), and these path descriptions must be taken further to computational instantiation (Klein, 2007a,b).
Anticipatory Thinking The importance of anticipation is widely acknowledged, in perception theory (e.g., top-down models of perception), human engineering and control theory (i.e., process control), and other literatures as well (Billings, 1996). Anticipatory thinking is more than prediction because people are preparing themselves for future events, not simply predicting what might happen. Anticipatory thinking includes active attention management—focusing attention on likely sources of critical information, reacting to trends, and apprehending the implications of combinations of events. Anticipatory thinking is typically aimed at low-probability, high-threat events, the ones where robustness is tested and adaptability and flexibility are most crucial to success. Available descriptions of anticipatory thinking must be taken further to computational instantiation (Klein and , 2011, 2007).
Coordination Robustness and adaptability of decision making pertains to team work as well as individual work. Members of teams must establish and maintain "common ground." When problems are discovered, teamwork hinges on mutual predictability and directability and other coordinative mechanisms. Attempts to build collaboration management agents (e.g., Allen and Ferguson, 2002) are predicated on identifying the rules that allow agents to be team players (Bradshaw, et al., 2004). Descriptions of paths through which humans achieve coordination in team decision making must be taken further toward computational instantiation (Christofferson and Woods, 2002; Klein, et al., 2004).

Whether one is describing the cognitive processes that characterize decision making or the software mechanisms that constitute a computational model, it is important to remember that in both cases we are dealing with a set of abstractions that we impose to facilitate explanation, description, and prediction. Thus, iterative exploration of the conceptual-computational is instructive of how to refine both the model and the theory (Warwick and Hutton, 2007). At each iteration, one must fix the level of abstraction at which there might be a correspondence between the conceptual (theoretical) model that the computational model. The modeling effort establishes a reciprocal relationship between the two models: The theoretical description informs the computational model and the computational model helps us explore aspects of the empirical description that remain under-specified. For instance, RPD does not assert that recognition involves comparing a list of cues to a memory record of all prior experiences. Conversely, the computational RPD model is mute concerning ways in which experts might perceptually integrate individual cues into patterns or into "chunks."

The effort to develop a computational model of Recognition-primed Decision Making made it clear that a critical step is the formation of sets of alternative assumptions that are necessary in order for implementation to proceed. Typically, "modeling" refers to a three-step process of increasing the scope of some existing model and then extending it in some way. The difficulty lies not in the development of novel algorithms, but rather in understanding what can, what must, and what should and can be represented (Warwick and Hutton, 2007).

If NDM models are indeed the most descriptive models we have of how real (proficient) people make real decisions then those NDM models must be formative of work methods and the technologies upon which the "real world" decision making relies. *Funding programs addressing specific issues in decision making, sensemaking, etc. might explicitly support attempts to develop computational instantiations of NDM conceptual models.*

Similarly, NDM-informed modeling should also be applied to the analysis of adversaries, who are themselves considered to be domain experts who engage in sensemaking and flexecution activities. A specific emerging challenge is to model and thereby anticipate the formulation of adversary intent in both real world and cyber domains, and anticipate how this is translated into action.

This includes modeling and anticipating adversary improvisation, adaptation and creativity in both real world and cyber domains. It includes anticipating and detecting the cues of emergent threat in complex settings.

Measuring Brittleness, Resilience, and Robustness

NDM has inspired theories of macrocognitive work that identify adaptivity and resilience as ideal goals (Hoffman and Woods, 2011; Woods, 2000). But how are such aspects of work systems to be measured? Robust decision making involves more than consistency in making good decisions, it involves identifying an option that will result in satisfying outcomes across the broadest swath of possible futures, making good decisions under circumstances where events are unfolding, problems are emergent, and stakes are high. More even than this, robust decision making includes a capability to change the way one makes decisions, in light of novelty and emergence.

The need for measures that illuminate features and phenomena at the "systems level" is widely recognized. Traditionally, human performance is gauged in terms of efficiency measures referred to as "HEAT" measures: hits, errors, accuracy and time (Hoffman, 2010). Such measures speak to the de-humanized economics of work systems, and are blind to other significant aspects of work systems. Is the work method learnable? Does it help workers achieve expertise? Does it motivate or demotivate workers? Are the tools understandable and usable? Are the humans and machines engaged in a genuine interdependence relationship in which they can make their intent and goals observable? (see Hoffman, Hancock and Bradshaw, 2010; Hoffman, et al., 2010; Klein, et al., 2004).

The concept of "resilience engineering" has gained significant traction in the engineering and computer science disciplines (Hollnagel, Woods and Leveson, 2006). It is now a topic for symposia on resilience in cyber systems, control systems, and communication systems. Recent funded research programs include calls for the development of technologies that manifest adaptive and resilient capacities. As we have seen for many concepts that make it to the front burner, resilience may be watered down and become a mere flavor of the month through overuse and uncritical use. That is, unless a methodology is forthcoming to specify ways in which resilience might actually be measured. So, what is resilience and how can it be measured in a way that enables the creation of human-centered technologies and macrocognitive work systems?

A number of different concepts of resilience have been discussed in the literature (Woods, 2015). One such meaning, which basically merges notions of resilience and adaptivity, is "robustness," the ability of a work system to maintain effectiveness across a range of tasks, situations, and conditions. A related concept is "flexibility," or the capacity to engage multiple paths to goals (Alberts and Hayes, 2003). Another meaning describes resilience as a form of "rebound" (see Woods, 2015), implying that the system's goals and methods have not fundamentally changed and the system gets "back on track" after it has experienced some sort of surprise. But this too is a notion that we would refer to as adaptivity rather than resilience.

- Adaptivity is the capacity of a system to achieve its goals despite the emergence of circumstances that push the system *toward* the boundaries of its competence envelope. The work system can employ multiple ways to succeed, or develop new ways to succeed, and can move seamlessly among them. The work system can reallocate and re-direct its resources to move away from the boundaries of its competence envelope and achieve its primary task goals.

- Resilience is the capacity to change as a result of circumstances that push the system *beyond* the boundaries of its competence envelope. The system may have to change some of its goals, procedures, resources, responsibilities components—any of its system-internal aspects. Because of those changes, the work system has a changed competence envelope. In effect, it becomes some other category of system (Woods and Branlat, 2011).

For measurement at the work system level the most important measures will be relativized, compound measures. So, for example, a measure of "number of adaptations" would have to be relativized on the assumption that the work system was able, in the scenario under scrutiny, to actually achieve its primary goals. If it was not successful, one have to consider that the work system was not very adaptive, no matter how many process changes were made.

In order for a macrocognitive work system to be adaptive and resilient the humans and the machine agents must work in a genuine interdependence relationship (Johnson, et al., 2014). To be an effective team player, human and machine agents must be able to adequately model the other participants' intents and actions. Human and machine team members must be mutually predictable. They must be directable. Agents must be able to make their status and intentions obvious to their teammates, and then the agents must be able to observe and interpret pertinent signals of the status and intentions of the other agents. Human and machine agents must be able to engage in goal negotiation. And finally, human and machine agents must be able to participate in the management of attention.

It is recommended that a research program explicitly and specifically focus on the challenges of creating usable and useful measures of adaptivity and resilience at the level of the macrocognitive work system. Effort along these lines could contribute significantly to theory, methodology, and the applications of the concepts of NDM and macrocognition.

Develop Means and Opportunity

The NDM community may not realize how much it can contribute. Even if individuals or groups within the community do realize what NDM has to offer, the community may currently not have the means or the opportunity to contribute. In simple terms, the community most likely has the latent motivation, but not the means or opportunity to address these challenges.

- <u>Communication</u>. There is a need for greatly enhanced two-way communication between challenge owners and the NDM community. The owners of these challenges (or at least, those charged with advancing solutions) need to communicate their needs better to the NDM community. The NDM community must be able to communicate its capabilities and potential benefits to decision makers and budget holders in terms of its methods, tools, outputs, and success stories; and to translate this into the potential contribution it might make towards addressing the challenges identified.

- <u>Span the Stakeholder Gaps</u>. For many large problems, there often exists a schism between those who hold (1) *Responsibility* (i.e. those who have been assigned, or have assumed, responsibility for addressing the problem); (2) *Authority* (i.e. those who can make things happen, usually the budget holders or those who can sign-off action); and (3) *Competency* (i.e. those who are technically capable of making a difference). To span these gaps, programs must be created in which all three are present and working together, otherwise effective progress cannot be made.

- <u>Make more Effective Use of Available Funding</u>. Recent cases of incidents that involved deficient human-system integration make it abundantly clear that NDM has not only "something to offer" but has scientific solution paths solve to significant challenges that confront the military. Although NDM researchers are intrinsically motivated to aid society, government, and the military, funding makes a significant contribution to motivation. The availability of funding is both a sign that challenge owners are serious about trying to solve the problems, and a sign that they believe that NDM research has the capability to contribute to the solution. However, whilst a piecemeal and scattergun approach to funding across a wide range of providers can increase coverage, and potentially yield increased innovation, a more focussed and sustained funding effort is likely to create the impetus for producing sustained and effective solutions. This suggests more coordination among the branches of the military (such as the effort of the Human Systems Community of Interest under the ASD/R&E) to avoid duplication of programs having the same problem sets and goals.

- Join the (solution) dots. It is highly unlikely that the capability for addressing the Challenges exists in one place, so in addition to 'joining the dots' to understand and make sense of the Challenges, we need to consider also 'joining the dots' to develop solutions. Existing pockets of expertise, potentially spread globally, must be joined-up and focussed synergistically on these problems. Collaboration, open information sharing and shared understanding, and close coupling of research and action will be required to drive forward initiatives that impact on these challenges.

- Coordinate the research effort across the NDM community. When funding is divided across multiple organizations, resultant research often feeds into the client organization in such a manner that there sometimes is nobody responsible for doing the 'joining-up'. More effort could be made to enable those in the NDM community to be more openly collaborative and less privately competitive on behalf of the challenge owners. This will include giving research organisations the big picture and strategy, allowing them to understand what others are doing and where they are heading, and (if IP challenges can be overcome) to share research products as they emerge (and not at the end of the research period, if at all).

- Promote continuity. For workers in the NDM community to feel like they have a real chance to make a dint against the Challenges, they require longitudinal projects involving multi-year funding, and (importantly) continuity of team members such that expertise can be developed, retained, exploited and effectively passed on to a next generation of STEM workers and scientists. Short-termism, piecemeal funding, and rapid staff turnover undermine any effort to make real progress against these challenges.

- Harness passion. Many within the NDM community are passionate about their work. This passion should be directed towards these big challenges, once people can see that they have an opportunity to make a real difference. Perhaps the broader NDM community should also be harnessed with regard to their view about how their capabilities align with the Challenges, and their ideas about how the Challenges may be tackled effectively (especially given that they work in a field concerned with naturalism).

- Navigate classification issues. Many of the Challenges problems involve highly sensitive subjects pertaining to national and global security. Addressing these successfully will require careful navigation of classification issues, such that expertise can be deployed effectively without inhibition, whilst limiting access to certain parts of the challenge space to appropriately cleared personnel. Such issues become more complex when working internationally. One path to a solution i for NDM researchers to engage their research skills in studies within the private sector, which has the same acute problems in cyber as does the government.

- Promote genuine relationships. When a sponsor prepares a statement of requirement for research organizations to bid, the sponsor naturally describes the presenting problem. That problem often involves expressing a desire to accomplish goals that go well beyond current scientific and methodological capabilities. One of the primary ways in which a healthy and productive collaboration can be established between sponsors and researchers is for sponsors to be more openly welcoming of "things they may not really want to hear." Although the need may have arisen from observations of specifics, the descriptions of requirements necessarily abstract away from detail, describing problems in terms of a particular conceptual jargon that the sponsor sees as appropriate. On winning the bid the research organization may then go in search of exemplars of the situations for detailed study, and those exemplars of course need to be relevant. It may be then that you realise that the abstract descriptions and jargon can represent a rather narrow bandwidth of communication for fixing the presenting problem. For effective data gathering, research requires a very specific focus, and this can make it difficult to 'cash out' the abstractions and jargon. The process should be openly and honestly negotiated. A big part of any research endeavour involves discovering what the problem actually is. Research involves uncovering that. What the problem really is, is often not what the sponsor thinks it is. Indeed, on can argue that for such topics as posed by the Emerging Challenges, the real scientific questions are never completely known at the outset, with and there is always 'room for interpretation.'

- Do not encourage proposals that are mere promissory notes. The term "User-Centered Design" has been with us for decades. So has "Work-Centered Design." So has "Work-Oriented Design." Indeed, a host of hyphenated designations have been published and proclaimed (reviewed in Hoffman, et al., 2002). Most of the design methods are getting at the same basic point—of what is really important about technology design. All of them decry everything that is "traditional," and proclaim to be qualitatively different from everything that has come before. All of them profess to do what no other design approach can do. We live in a competitive climate in which everyone—whether in the private sector or in the academic sector—has to make their work seem special and assert themselves as "uniquely qualified." Everyone promises to perform miracles, proclaiming in present tense the capabilities of a software system that it has not even yet been built. Individual researchers in the computational and decision sciences would serve their sciences, and their communities of practice, by being more honest, and fair.

- Conceptualise requirements at many levels of abstraction considering both need and opportunity. What is unknown at the outset of a research endeavour is exactly how much of an activity can be re-engineered using a hypothetical new technology, and in what ways. It may be that reducing the user-costs of interaction, or enabling collaboration can involve relatively surface-level changes to current practice or it may mean that there is an opportunity for a more radical redesign to fundamentally change the way that higher-level goals are achieved. The open question is what at what level re-engineering is and can be done and this is a question not just of what the activity is but what opportunities are presented by new interventions.

- Encourage honesty in management and in reporting. History shows that the vast majority of scientific experiments are failures, yet we live in a climate in which all experiments are successes, everything deserves to be published, and all R&D programs are programmatic successes. Constraints of exigency and economics are such that research is designed by spread sheet and managed by schedule. While there are need for schedules and milestones, it should be remembered that genuine science does not always proceed by timetable. There are fits and starts. There is back-peddling. Our nation's scientific infrastructure would benefit by more open acknowledgement of and accommodation to these aspects of genuine science, if only to reinforce rather than dampen the intrinsic motivation of researchers to do good science and to have positive impact.

Speak in Many Tongues but Collaborate in One

The "Emerging challenges" are large, complex and multifaceted; and may well require large, complex and multifaceted solutions. The challenges are far broader in scope than the field of NDM is able to cope with alone, so a cross-disciplinary approach, of which NDM is part, likely stands the best chance of yielding impactful solutions. The NDM community has long recognized that their focus is on applied problems that call for multidisciplinary methodologies. NDM researchers engage in cognitive ethnography, though cognitive ethnography is a historically different community of practice, with different focus points for its research. NDM researchers engage in research that might be thought of as industrial/organizational psychology, though I/O psychology is a historically different discipline as well as a different community of practice. NDM research is related to human factors psychology/human factors engineering, and also to cognitive systems engineering. These disciplines are historically associated more with the design and implementation of technologies whereas NDM is historically associated more with psychological experimentation and cognitive field research. NDM sees itself as feeding into human factors and cognitive systems engineering, and much NDM research is presented at human factors meetings. NDM researchers engage in technology design, though design is itself the focus of a number of different communities of practice in a number of different engineering disciplines.

Almost all of the specific problems inherent in each of the Challenge are generic, enduring, and cut across all the challenges identified. Rather than stove-pipe the exploration of solutions by challenge type or domain, challenges should be tacked in a cross-domain manner, such that investment made in solution generation can be leveraged across all domains and problem sets.

Different disciplines bring with them different constraints on solutions and different conceptual languages for describing problems and solution opportunities; they speak in many tongues. One community of practice might prefer to speak of "human behavior" whereas another might prefer to speak of "human activity," claiming that the otherwise innocuous word "behavior" actually carries considerable historical baggage. One community might decry reference to "automation" on the argument that so-called autonomous technologies are in fact never autonomous; whereas another community might reply that "Well, of course machines are not completely autonomous."

One community might use the word "system" to refer to the humans-machines work system whereas another might reserve the word "system" to refer to just the technology. And so forth. (See Hoffman and Hancock, 2014a,b).

Every discipline necessarily develops a vocabulary to encode its knowledge and values, about methods, cognitive phenomena, design patterns, or ethical principles, or legal principles. In the context of project work these can be shared rather fleetingly, appearing merely as bullet points of jargon on PowerPoint presentations. The challenge here is to engage in a deep way with each collaborator's view of the world. This necessarily involves more than the integration of disciplines. The formula of "Add three psychologists and stir well." will not work. The issue is one of learning and the filling of responsibility gaps, not an issue of cross-disciplinary communication. Understanding has to be more than simply a surface level appreciation. Project teams coordinate best when they grow together, and can think using each other's language. It is a good sign when members of interdisciplinary groups start to adopt each other's language in their articulation of problems and solutions. Recent collaborative successes attest to this. (See for instance Johnson, et al., 2015.)

The perpetual reinvention of wheels, and re-discovery of lessons learned is the main result of the super-fragmentation (or hyper-specialization) of disciplines, the overwhelming proliferation of specialized journals, and the proliferation of diverse communities of practice. Thus, cognitive ethnographers might publish a paper proclaiming a new method for task analysis, to which human factors psychologists might say "Oh, we invented that sort of thing decades ago." Computer scientists might emphasize the importance of situational awareness when experimental psychologists would say that basic notion can be found in the literature of psychology dating back to the 1880s. Human Factors psychologists might begin programmes of experiments on team composition, to which I/O psychologists would respond that they have been doing that for decades. And so forth. Reinvention/rediscovery can be regarded as verification or validation of good ideas. But the lack of general awareness across specializations and across communities of practice, and the general ignorance of history, are impediments to collegial, collaborative progress on the really important problems.

The consequences of the lack of historical scholarship and training cannot be over-emphasized. Indeed, today many undergraduate and graduate programs in experimental psychology, even in elite schools, never require students to take history of psychology courses.

Unfortunately, little can be done about this super-fragmentation. If anything, the situation will get progressively worse. But there is one thing that the NDM community can, indeed must do, and that is drop its own historical baggage. The designation "Naturalistic Decision Making" reflected the origins of the community of practice in the 1980s, as a reaction to the dominant normative-rationalist view of decision making that dominated economics for many years. For decades now, NDM-ers have to always explain that NDM is not just about decision making: it is about processes including sensemaking, replanning, collaborating, and others. Ever since the Seventh International Conference on NDM, there has been a growing recognition that the name of the community has to change. This need was emphasized at the 2015 meeting in the Keynote presentation by David Woods. Thus, it is proposed that the next NDM meeting be titled *"The 2017 International Conference on Macrocognition: Expanding the Horizons of Naturalistic*

Decision Making." The title is intended to commence a re-branding, while the subtitle is meant to help insure continuity. As long as NDM-ers have to always explain that NDM is not just about decision making, they might as well explain the concept of macrocognition.

There is clear historical precedent for the concept of macrocognition. Discussions of that history, and of macrocognition as the genuine foundation for NDM, appear in Hoffman, Klein and Schraagen, 2007; Hoffman and Woods, 2011; Klein and Hoffman, 2008; Klein, Moon and Hoffman, 2006a,b; Schraagen, Klein, and Hoffman, 2008. Especially pertinent review papers are:

Hoffman, R.R. (2010). Some challenges for macrocognitive measurement. In E. Patterson and J. Miler (Eds.), *Macrocognition metrics and scenarios: Design and evaluation for real-world teams* (pp. 11-28). London: Ashgate.

Hoffman, R. R., and McNeese, M. (2009). A history for macrocognition. *Journal of Cognitive Engineering and Decision Making, 3,* 97-110.

Klein, G., Ross, K.G., Moon, B.M., Klein, D.E., Hoffman, R.R., and Hollnagel, E. (May/June, 2003). Macrocognition. *IEEE: Intelligent Systems*, pp. 81-85.

The NDM community, and indeed many other communities of practice, would benefit from a better understanding of how to effectively conduct interdisciplinary team science. The Emerging Challenges are unlikely to be solved quickly. Approaches that adopt a more longitudinal view are required, that have substantial funding across multiple years, and that employ stable cross-organization multidisciplinary teams over such periods. Further, there is unlikely to be one 'solution' to any of these challenges that can be prescribed in advance of implementation. Rather, an incremental, experientially-driven process of learning by doing will be required, that is agile and able to adapt readily to hard-won lessons as they arise.

Designer-Centered Design and Procurement Policy

The benefits of computerization for electronic management of complex and distributed health-care information would, at first glance, seem irrefutable. Challenger, Clegg and Shepherd (2013) described the experience of the UK National Health Service with an electronic healthcare record known as the NHS Care Records Service. This system was developed to manage medical records for all patients in the UK National Health Service. The goal was to ensure that every patient's healthcare information would be integrated into a single record that could be accessed at any location within the UK National Health Service. Despite the many obvious benefits of computerization, there was widespread evidence of diverse problems; a lack of compatibility with clinical practice, incomplete and inaccurate information, a restrictive data entry strategy, and an electronic-notes function that increased the cognitive work associated with taking a patient history, to name just a few. There was little evidence that the anticipated benefits were realised. A similar situation obtains for the push to electronic health care records in the U.S.

From the perspective of NDM. the sorts of problems experienced with electronic health records are unsurprising. The design strategy is essentially a technology-push driven by a political agenda. A limited subset of stakeholders (in health care, managers and a small selection of medical staff) envisioned a response to a political mandate. In effect, their response is little more than a recommendation to proceed with a particular type of technological solution (see Neville, et al., 2008). From there, the design, development, and deployment of the technological solution was in the hands of the technologists. Rarely do those technologists have anything more than a superficial understanding of the work their system is intended to support.

Despite the recurring evidence that systems designed, developed and deployed in this manner fail to satisfy the need, no one appears to attribute the failures to the strategy and the policy behind it. No one outside of cognitive engineering, it seems, can conceive of another way. The idea that a technological development should be driven by the need to better support the work practices and the work goals of the diverse stakeholders within the system (rather than by the desire to computerize work practices) is not one that comes naturally to those who direct the development of these systems.

This is one of the major challenges facing our society and nation and it is one that, if resolved, would enable NDM to make an enormous contribution. Conceptually at least, the NDM strategy of designing from a thorough understanding of the work of all stakeholders is straightforward. A strength of NDM research is the emphasis on human capabilities. Current funding programs addressing the Emerging Challenges tend to focus on technological solutions, with the assumption that humans will effortlessly adapt and acquire the skills needed to integrate, monitor, maintain, and collaborate with new technologies. If funding agencies and research leaders were more open to exploring strategies for supporting humans and strengthening human-technology interdependence, more NDM researchers might get involved in research on topics related to the Emerging Challenges.

Few examples exist in which a macrocognitive lens of NDM has been applied at the policy level, or in articulating the overarching architecture for a complex work system. Generally, other disciplines are called on to design and create these high-level frameworks, and NDM researchers get involved as sticky and particular problems emerge. Thus, NDM models and methods are grounded in exploring specific incidents in the context of a particular domain. This often leads to very effective interventions that help overcome important barriers. Applying these methods and models across the multiple domains that must work together to solve these complex emerging challenges would require adaptation of existing methods and perhaps new models for representing cognitive work with clear implications for policy and large-scale system design.

References

Alberts, D.S. and Hayes, R.E. (2003). Power to the Edge: Command and Control in the Information Age. C2 Research Program Publications.

Annett, J. (2000). Theoretical and pragmatic influences on task analysis methods. In J.-M. Schraagen, S. Chipman, and V. Shalin (Eds.) *Cognitive task analysis.* Mahwah, NJ: Erlbaum.Beach, L. R. (1990). *Image theory: Decision-making in personal and organizational contexts*. Chichester, UK: Wiley.

Beach, L. R. (1993). Image theory: Personal and organizational decisions. In Klein, G. A., Orasanu, J., Calderwood, R., and Zsambok, C. E. (Eds.), *Decision making in action: Models and methods* (pp. 148-157). Norwood, NJ: Ablex.

Beach, L. R., Chi, M. T. H., Klein, G., Smith, P., and Vicente, K. (1997). Naturalistic decision making research and related lines. In C. E. Zsambok and G. Klein (Eds.), *Naturalistic decision making* (pp. 29-36). Mahwah, NJ; Erlbaum.

Beach, L. R., and Lipshitz, R. (1993). Why classical decision theory is an inappropriate standard for evaluating and aiding most human decision making. In Klein, G. A., Orasanu, J., Calderwood, R., and Zsambok, C. E. (Eds.), *Decision making in action: Models and methods* (pp. 21-35). Norwood, NJ: Ablex.

Beach, L. R., and Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review, 3*, 439-449.

Billings, C. E. (1996). *Aviation automation: The search for a human-centered approach.* Mahwah, NJ: Lawrence Erlbaum Associates.

Boring, E. G. (1950). *A history of experimental psychology.* Boston: DC Heath.

Calderwood, R., Crandall, B., and Klein, G. (1987). "Expert and novice fireground command decisions." Report MDA903-85-C-0327. U.S. Army Research Institute, Alexandria VA.

Challenger, R., Clegg, C.W. and Shepherd, C. (2013). Function allocation in complex systems: reframing an old problem, *Ergonomics, 56,* 1051-1069
DOI: 10.1080/00140139.2013.790482

Brehmer, B. (1991). Distributed decision making: some notes on the literature. In J. Rasmussen, B. Brehmer and J. Leplat (Eds.), Distrib*uted decision making: Cognitive models for cooperative work* . Chichester, England: John Wiley.

Brooking, A. (1999). *Corporate memory: Strategies for knowledge management.* London: International Thomson Business Press.

Cacciabue, P. C., and Hollnagel, E. (1995). Simulation of cognition: Applications. In J. M. Hoc, P. C., Cacciabue, and E. Hollnagel (Eds.), *Expertise and technology: Cognition and human-computer cooperation* (pp. 55-73). Hillsdale, NJ: Lawrence Erlbaum Associates.

Calderwood, R., Crandall, B., and Klein, G. (1987). "Expert and novice fireground command decisions." Report MDA903-85-C-0327. U.S. Army Research Institute, Alexandria VA.

Chandrasekaran, B. (2005). *From optimal to robust COAs: challenges in providing integrated decision support for simulation-based COA planning.* Columbus, OH: Laboratory for AI Research, The Ohio State University.

Chandrasekaran, B., and Goldman, M. (2007). Exploring robustness of plans for simulation-based course of action planning: A framework and an example. *Proceedings of the IEEE Symposium on Computational Intelligence in Multicriteria Decision Making*, Honolulu, HI.

Chi, M. T. H, Feltovich, P. J., and Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121-152.

Christensen-Szalanski, J. J. (1993). A comment on applying experimental findings of cognitive biases to naturalistic environments. In G. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok, C. E. (Eds.) *Decision making in action: Models and methods* (pp. 252-264). Norwood, NJ: Ablex.

Chow, R. (2007). Personal communication. Human Systems Integration Section, Defence Research and Development, Toronto, Canada.

Cohen, M. S. (1993). The naturalistic basis of decision biases. In Klein, G. A., Orasanu, J., Calderwood, R., and Zsambok, C. E. (Eds.), *Decision making in action: Models and methods* (pp. 51-102). Norwood, NJ: Ablex.

Cohen, M. S., Freeman, J. T., and Wolf, S. (1996). Metarecognition in time-stressed decision making: Recognizing, critiquing, and correcting. *Human Factors, 38*, 206-219.

Crandall, B. (1989). A comparative study of think-aloud and critical decision knowledge elicitation methods. *SIGART Newsletter*, No. 108, pp. 144-146. New York: Special Interest Group on Artificial Intelligence, Association for Computing Machinery.

Crandall, B., and Calderwood, R. (1989). "Clinical assessment skills of experienced neonatal intensive care nurses". Report, Klein Associates, Yellow Springs OH.

Crandall, B., and Gamblian, V. (1991). "Guide to early sepsis assessment in the NICU". Report, Klein Associates, Yellow Springs OH.

Crandall, B., and Getchell-Reiter, K. (1993). Critical decision method: A technique for eliciting concrete assessment indicators from the intuition of NICU nurses. *Advances in Nursing Science, 16*, 42-51.

Crandall, B. W., and Klein, G. (1987a). "Key components of MIS performance". Report, Klein Associates, Yellow Springs, OH.

Crandall, B. W., and Klein, G. (1987b). "Critical cues for MI andcardiogenic shock symptoms". Report, Klein Associates, Yellow Springs OH.

Crandall, B., Klein, G., and Hoffman R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. Cambridge, MA: MIT Press.

Dawes, R. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*, 571-582.

Dawes, R. M., and Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81*, 95-106.

De Garmo, C. (1895). *Herbart and the Herbartians.* New York: Scribner.

De Groot, A. D. (1945). *Het denken van der schaker*. Ph.D. Thesis, University of Amsterdam.

Dember, W. N., and Warm, J. S. (1979). *Psychology of perception* (2nd ed.). New York: Holt, Rinehart and Winston.

Drury, C. G., Paramore, B., Van Cott, H., Grey, S. M., and Corlett, E. N. (1987). Task analysis. In G. Salvendy (Ed.), *Handbook of human factors* (pp. 370-401). New York: John Wiley.

Duncker, K. (1945). On problem solving. *Psychological Monographs*, No. 58, 1-113 (Whole No. 270) (L. S. Lees, Trans.).

Ebbinghaus, H. (1908). *Psychology: An elementary text* (M. Meyer, Trtans.). Boston: D. C. Heath and Co.

Edwards, W. (1965a). Optimal strategies for seeking information: Models for statistics, choice reaction time, and human information processing. *Journal of Mathematical Psychology,* 2, 312-329.

Einhorn, H. J. (1974). Expert judgment: Some necessary conditions and an example. *Journal of Applied Psychology, 59*, 562-571.
Endsley, M. R. (1988). Situation Awareness Global Assessment Technique (SAGAT). In *Proceedings of the National Aerospace and Electronics Conference (NAECON)* (pp. 789-795). New York: IEEE.
Endsley, M. R. (1990). Predictive utility of an objective measure of situation awareness. In *Proceedings of the Human Factors and Ergonomics Society 34th Annual Meeting* (pp. 41-45). Santa Monica, CA: Human Factors Society.
Endsley, M. R. (1993). A survey of situation awareness requirements in air-to-air combat fighters. *International Journal of Aviation Psychology*, *3*(2), 157-168.
Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human Factors*, *37*, 65-84.
Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*, 32-64.
Endsley, M. R. (1997). The role of situation awareness in naturalistic decision making. In C. E. Zsambok and G. Klein (Eds.), *Naturalistic decision making* (pp. 269-284). Mahwah, NJ; ´Erlbaum.
Endsley, M., Bolte, B., and Jones, D. G. (2003). *Designing for situation awareness*. New York: Taylor and Francis.
Endsley, M. R., and Garland, D. J. (Eds.). (2000). *Situation awareness analysis and measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.
Endsley, A., and Hoffman, R. (November/December, 2002) The Sacagawea Principle. *IEEE Intelligent Systems,* pp. 80-85.
Ericsson, K. A., and Smith, J. (Eds.) (1991a). *Toward a general theory of expertise*. Cambridge: Cambridge University Press.
Eskridge. T. C., Still, D. and Hoffman, R.R. (2014, July/August). Principles for Human-Centered Interaction Design, Part 1: Performative Systems. *IEEE Intelligent Systems,* pp. 88-94.
Fan, X., McNeese, M., and Yen, J. (2010). NDM-based cognitive agents for supporting decision-making teams. *Human-Computer Interaction, 25,* 95-234.
Feltovich, P.J., Coulson, R.L., and Spiro R.J. (2001). Learners' (mis)undestanding of important and difficult concepts: A challenge to smart machines in education. In K. D. Forbus & P. J. Feltovich (Eds.), *Smart machines in education* (pp. 349-375). Menlo Park, CA: AAAI/MIT Press.
Feltovich, P.J., Hoffman, R.R., and Woods, D. (May/June 2004). Keeping it too simple: How the reductive tendency affects cognitive engineering. *IEEE Intelligent Systems*, pp. 90-95.
Feltovich, P.J., Spiro, R.J., and Coulson, R.L. (1997). Issues of expert flexibility in contexts characterized by complexity and change. In Feltovich, P.J., Ford, K., and Hoffman, R.R. (Eds.), *Expertise in Context,* (pp. 125-146). Menlo, CA: AAAI Press/MIT Press.
Fischhoff, B. (1989). Eliciting knowledge for analytical representation. *IEEE Transactions on Systems, Man, and Cybernetics, 19*, 448-461.
Fischhoff, B., Goitein, B., and Shapira, Z. (1982). The experienced utility of expected utility approaches. In N. Feather (Ed.), *Expectations and actions: Expectancy value models in psychology*. Hillsdale, NJ: Erlbaum.
Fischhoff, B., Slovic, P., and Lichtenstein, S. (1979). Subjective sensitivity analysis. *Organizational Behavior and Human Performance, 23*, 339-359.

Flach, J., and Hoffman, R. R. (January-February 2003). The limitations of limitations. *IEEE Intelligent Systems.* (Pp. 94-97).

Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin, 51*, 327-358.

Flin, R., Salas, E., Strub, M., and Martin, L. (Eds.) (1997). *Decision making under stress: Emerging themes and applications*. Aldershot, UK: Ashgate Publishing.

Gettys, C. F., Fisher, S. D., and Mehle, T. (1978). "Hypothesis generation and plausibility assessment." Technical Report No. 15-10-78. Decision Process Laboratory, university of Oklahoma, Norman OK.

Goldberg, L. R. (1965). Diagnosticians vs. Diagnostic Signs: The Diagnosis of Psychosis vs. Neurosis from the MMPI. *Psychological Monographs, 70*(9), 1-28.

Gordon, S. E., and Gill, R. T. (1997). Cognitive task analysis. In C. Zsambok and G. Klein (Eds.), *About naturalistic decision making* (pp. 131-149). Mahwah, NJ: Erlbaum.

Grove, W. M., and Meehl, P. E. (1996). Comparative efficiency of formal (mechanical, algorithmic) and informal (subjective, impressionistic) prediction procedures: The clinical/statistical controversy. *Psychology, Public Policy, and Law, 2*(2), 293-323.

Hammond, K. R. (1966). Clinical inference in nursing, II: A psychologist's viewpoint. *Nursing Research, 15*, 27-38.

Hammond, K. R. (1993). Naturalistic decision making from a Brunswikian viewpoint: Past, present, future. In Klein, G. A., Orasanu, J., Calderwood, R., and Zsambok, C. E. (Eds.), *Decision making in action: Models and methods* (pp. 205-227). Norwood, NJ: Ablex.

Hammond, K. R. Hamm, R. M. Grassia, J., and Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics, 17*, 753-770.

Hammond, K. R., Kelly, K. J., Schneider, R. J., and Vancini, M. (1966). Clinical inference in nursing: Information units used. *Nursing Research, 15*, 236-243.

Hammond, K. R., McClelland, G. H., and Mumpower, J. (1980). *Human judgment and decision making: Theories, methods, and procedures.* New York: Praeger.

Hancock, P.A. (1986). The effect of skill on performance under an environmental stressor. *Aviation, Space, and Environmental Medicine, 57, 59-64.*

Hancock, P.A. (2007). On the process of automation transition in multi-task human-machine systems. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Humans and Systems, 37, 586-598.*

Hancock, P.A., Szalma, J.L., and Oron-Gilad, T. (2005). Time, emotion, and the limits to human information processing. In: D. McBride and D. Schmorrow (Eds.). *Quantifying human information processing.* (pp. 157-175), Lexington Books: Boulder, CO.

Hancock, P.A., Weaver, J.L., and Parasuraman, R. (2002). Sans subjectivity, Ergonomics is Engineering. *Ergonomics, 45,* 991-994.

Hewitt, K. (1983) *Interpretations of calamity: From the viewpoint of human ecology*. London: Allen and Unwin.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*, 411-428.

Hoc, J.-M., Cacciabue, P. C., and Hollnagel, E. (Eds.) (1996). *Expertise and technology: Cognition and human-computer cooperation.* Mahwah, NJ: Erlbaum.

Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin, 57,* 116-131.

Hoffman, R.R. (Ed.) (2007). *Expertise out of context: Proceedings of the Sixth International Conference on Naturalistic Decision Making*. Boca Raton FL: Taylor and Francis.

Hoffman, R. R. (1987b, Summer). The problem of extracting the knowledge of experts from the perspective of experimental psychology. *The AI Magazine, 8*, 53-67.

Hoffman, R. R. (1991a). Human factors psychology in the support of forecasting: The design of advanced meteorological workstations. *Weather and Forecasting, 6*, 98-110.

Hoffman, R. R. (Ed.) (1992b). *The psychology of expertise: Cognitive research and empirical AI.* Mahwah, NJ: Erlbaum.

Hoffman, R. R. (1995a). "A Review of Naturalistic Decision Making on the Critical Decision Method of Knowledge Elicitation and the Recognition-Priming Model of Decision-making, with a Focus on Implications for Military Proficiency." Report to Epistemics, Ltd., under a contract from the Defense Research Agency, Ministry of Defense, United Kingdom.

Hoffman, R. R. (1998b). How can expertise be defined? Implications of research from cognitive psychology. In R. Williams and J. Fleck (Eds.), *Exploring expertise: Issues and perspectives*. London: Macmillan.

Hoffman, R.R. (2010). Some challenges for macrocognitive measurement. In E. Patterson and J. Miler (Eds.), *Macrocognition metrics and scenarios: Design and evaluation for real-world teams* (pp. 11-28). London: Ashgate.

Hoffman, R. R., Coffey, J. W., Carnot, M. J., and Novak, J. D. (2002, September). "An empirical comparison of methods for eliciting and modeling expert knowledge." In *Proceedings of the 46th Meeting of the Human Factors and Ergonomics Society* (pp. 482-486). Santa Monica, CA: Human Factors and Ergonomics Society.

Hoffman, R. R., Coffey, J. W., and Ford, K. M. (2000). "A Case Study in the Research Paradigm of Human-Centered Computing: Local Expertise in Weather Forecasting." Report on the Contract, "Human-Centered System Prototype," National Technology Alliance.

Hoffman, R. R., Coffey, J. W., Ford, K. M., and Novak, J. D. (2006). A method for eliciting, preserving, and sharing the knowledge of forecasters. *Weather and Forecasting, 21*, 416–428.

Hoffman, R. R., Crandall, B., and Shadbolt, N. (1998). A case study in cognitive task analysis methodology: The Critical Decision Method for the elicitation of expert knowledge. *Human Factors*, *40*, 254-276.

Hoffman, R. R., and Feltovich, P. J., and Ford, K. M. (1997). A general framework for conceiving of expertise and expert systems in context. In P. J. Feltovich, K. M. Ford, and R. R. Hoffman (Eds.).*Expertise in context: human and machine* (pp. 543-580). Cambridge, MA: MIT Press/AAAI Books.

Hoffman, R.R., Feltovich, P.J., Ford, K. M., Woods, D. D., Klein, G. and Feltovich, A. (July/August 2002). A rose by any other name... would probably be given an acronym. *IEEE: Intelligent Systems*, 72-80.

Hoffman, R. R., and Fiore, S. M. (2007, May/June). Perceptual (Re)learing: A leverage point for Human-Centered Computing. *IEEE Intelligent Systems,* pp. 79-83.

Hoffman, R.R. and Hancock, P.A. (2014a). (More) Words Matter. *Bulletin of the Human Factors and Ergonomics Society*, *57* (10) [http://www.hfes.org/web/HFESBulletin/oct2014morewords.html]

Hoffman, R.R. and Hancock, P.A. (2014b). Words Matter. *Bulletin of the Human Factors and Ergonomics Society*, *57* (8)

[http://www.hfes.org/web/HFESBulletin/aug2014wordsmatter.html]
Hoffman, R.R., Hancock, P.A., and Bradshaw, J.M. (2010, November/December). Metrics, metrics, metrics, Part 2: Universal Metrics? *IEEE Intelligent Systems*, pp. 93-97.
Hoffman, R.R. and Hanes, L.F. (July-August 2003). The boiled frog problem. *IEEE Intelligent Systems,* pp. 68-71.
Hoffman, R.R., Klein, G., and Laughery, K. R. (January/February 2002). The state of cognitive systems engineering. *IEEE: Intelligent Systems*, pp. 73-75.
Hoffman, R. R., Klein, G., and Schraagen, J. M. (2007). The macrocognition framework of Naturalistic Decision Making. In J. M. Schraagen (Ed.), *Macrocognition*. London: Ashgate.
Hoffman, R. R., and Lintern, G. (2006). Eliciting and representing the knowledge of experts. In K. A. Ericsson, N. Charness, P. Feltovich, and R. Hoffman, (Eds.) *Cambridge handbook on expertise and expert performance* (pp. 203-222). New York: Cambridge University Press.
Hoffman, R. R., Lintern, G., and Eitelman, S. (March/April 2004). The Janus Principle. IEEE: Intelligent Systems, pp. 78-80.
Hoffman, R.R., Marx, M., Amin, R., and McDermott, P. (2010). Measurement for evaluating the learnability and resilience of methods of cognitive work. *Theoretical Issues in Ergonomic Science*. Published online at iFirst, DOI: 10.1080/14639220903386757.
Hoffman, R. R., and McNeese, M. (2009). A history for macrocognition. *Journal of Cognitive Engineering and Decision Making, 3,* 97-110.
Hoffman, R.R. and Militello, L.G. (2008). *Perspectives on Cognitive Task Analysis*: *Historical Origins and Modern Communities of Practice*. Boca Raton, FL: CRC Press/Taylor and Francis.
Hoffman, R. R., Shadbolt, N. R., Burton, A. M., and Klein, G. (1995). Eliciting knowledge from experts: A methodological analysis. *Organizational Behavior and Human Decision Processes*, *62*, 129-158.
Hoffman , R.R., and Ward, P.(2015, in press). Intelligent Mentoring systems? *IEEE: Intelligent systems.*
Hoffman, R.R., Ward, P., DiBello, L., Feltovich, P.J., Fiore, S.M., and Andrews, D. (2014). *Accelerated Expertise: Training for High Proficiency in a Complex World.* Boca Raton, FL: Taylor and Francis/CRC Press.
Hoffman, R. R., and Woods, D. D. (2000). Studying cognitive systems in context. *Human Factors, 42,* 1-7.
Hoffman, R. R., and Woods, D. D. (January/February, 2005). Steps toward a theory of complex and cognitive systems. *IEEE: Intelligent Systems,* pp. 76-79.
Hoffman, R.R. and Woods, D.D. (2011, November/December). Beyond Simon's slice: Five fundamental tradeoffs that bound the performance of macrocognitive work systems. *IEEE: Intelligent Systems*, pp. 67-71.
Hoffman, R. R., and Yates, J. F. (July/August, 2005). Decision(?)-Making(?). *IEEE: Intelligent Systems*, pp. 22-29.
Hoffman, R.R., Ziebell, D., Fiore, S.M. and Becerra-Fernandez, I. (2008, May/June). Knowledge management revisited. *IEEE: Intelligent Systems*, pp. 84-88.
Hollnagel, E. (1982). "Cognitive Task Analysis." Draft Report in a memo to David Woods, Institutt for Atomenergi, Halden, Sweden.

Hollnagel, E. (2009). *The ETTO Principle: Efficiency-thoroughness trade-off: Why things that go right sometimes go wrong.* Farmham, UK: Ashgate.

Hollnagel, E., and Cacciabue, P. C. (1999). Cognition, technology, and work: An introduction. *Cognition, Technology, and Work, 1,* 1-6.

Hollnagel, E., Hoc, J.-M., and Cacciabue, P. C. (1996). Expertise and technology: "I have a feeling we're not in Kansas anymore." In J.-M. Hoc, P. C. Cacciabue, and E. Hollnagel, E. (Eds.) *Expertise and technology: Cognition and human-computer cooperation* (pp. 279-286). Mahwah, NJ: Erlbaum.

Hollnagel, E., Pedersen, O. M. and Rasmussen, J. (1981). "Notes on Human Performance Analysis." Report Riso-M-2285), Riso National Laboratory, Roskilde, Denmark.

Hollnagel, E., and Woods, D. D. (1983). Cognitive systems engineering: New wine in new bottles. *International Journal of Man-Machine Studies, 18,* 583-600.

Hollnagel, E., and Woods, D. D. (2006). *Joint cognitive systems: Foundations of cognitive systems engineering.* Boca Raton, FL: Taylor and Francis.

Hollnagel, E.,Woods D.D., and Leveson, N. (Eds.) (2006). *Resilience engineering: Concepts and Precepts.* London: Ashgate.

Howell, W. C. (1984). "Task influences in the analytic-intuitive approach to decision making." Report, Contract No. N00014-82-001. Office of Naval Research, Bethesda, MD.

Isenberg, D. J. (1984). How senior managers think. *Harvard Business Review, 6,* 80.

Stewart, T. R., Roebber, P. J., and Bosart, L. F. (1997). The importance of the task in analyzing expert judgment. *Organizational Behavior and Human Decision Processes, 69*(3), 205-219.

Janis, I. L. and Mann, L. (1977). *Decision-making: A psychological analysis of conflict, choice, and commitment.* New York: The Free Press.

Johnson, M., Bradshaw, J.M., Hoffman, R.R., Feltovich, P.J. and Woods, D.D. (November/December 2014). Seven cardinal virtues of human-machine teamwork. *IEEE Intelligent Systems*, pp. 74-79.

Kaempf, G. L., Wolf, S., Thordsen, M. L., and Klein, G. (1992). Decision making in the AEGIS combat information center (Contract N66001-90-C-6023 for the Naval Command, Control and Ocean Surveillance Center). Fairborn, OH: Klein Associates Inc.

Kahneman, D., Slovic, P., and Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Kahneman, D., and Tversky, A. (Eds.) (2000). *Choices, values, and frames.* Cambridge: Cambridge University Press.

Kirwan, B., and Ainsworth, L. K. (Eds.) (1992). *A guide to task analysis.* London: Taylor and Francis.

Kassirer, J. P., Kuipers, B. J., and Gorry, G. A. (1982). Toward a theory of clinical expertise. *American Journal of Medicine, 73,* 251-259.

Klein, G. (1987). Applications of analogical reasoning. *Metaphor and Symbolic Activity, 2,* 201-218.

Klein, G. (1989). *Utility of the critical decision method for eliciting knowledge from expert C debuggers.* Report on Purchase Order No. 339404, ATandT Bell Laboratories. Klein Associates, Inc., Yellow Springs, OH.

Klein, G. (1989). Recognition-primed decisions. In W. B. Rouse (Ed.), *Advances in Man-Machine Research, Volume 5.* JAI Press Greenwich, CT.

Klein, G. (1992). Using knowledge engineering to preserve corporate memory. In R. R. Hoffman (Ed.), *The psychology of expertise: Cognitive research and empirical AI* (pp. 170-190). Mahwah, NJ: Erlbaum.

Klein, G. (1993a). The recognition-primed decision model: Looking back, looking forward. In. C. Zsambok and G. Klein (Eds.), *Naturalistic decision making* (pp. 285-292). Mahwah, NJ: Erlbaum.

Klein, G. (1993c). *Naturalistic decision making--Implications for design.* Technical Report (Contract No. SOAR 93-01, Crew System Ergonomics Information Analysis Center, Wright Patterson Air Force Base, OH). Klein Associates Inc., Fairborn, OH.

Klein, G. (1995). The value added by cognitive task analysis. *Proceedings of the 39th annual Human Factors and Ergonomics Society Meeting* (pp. 530-533). Santa Monica, CA: Human Factors and Ergonomics Society.

Klein, G. (1997). The recognition-primed decision (RPD) model: Looking back, looking fowrard. In C. E. Zsambok and G. Klein (Eds.), *Naturalistic decision making* (pp. 285-292). Mahwah, NJ; Erlbaum.

Klein, G. (1998). *Sources of power*. Cambridge, MA: MIT Press.

Klein, G. (2003). *Intuition at work.* New York: Doubleday.

Klein, G., and Brezovic, C. P. (1986). Design engineers and the design process:Decision strategies and human factors literature. *Proceedings of the 30th Annual Meeting of the Human Factors Society* (pp. 771-775). Santa Monica, CA: Human Factors Society.

Klein, G., Calderwood, R., and Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground, *Proceedings of the 30th Annual Meeting of the Human Factors Society* (pp. 576-580). Santa Monica, CA: Human Factors Society.

Klein, G., Calderwood, R., and MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*, 462-472.

Klein, G., Feltovich, P. J., Bradshaw, J.M., and Woods, D.D. (2006). Common ground and coordination in joint activity. In W.R. Rouse and K.B. Boff (Eds.), *Organizational Simulation*. New York: Wiley.

Klein, G., and Hoffman, R. R. (2007). The use of cognitive task analysis methods to capture mental models, In J.M Schraagen (Ed.), *Proceedings of the Seventh International Conference on Naturalistic Decision Making*. London: Ashgate Publishing.

Klein, G. and Hoffman, R. R. (2008). Macrocognition, mental models, and cognitive task analysis methodology. In J. M. Schraagen, L. G. Militello, T. Ormerod and R. Lipshitz (Eds.), *Naturalistic decision making and macrocognition* (pp. 57-80). Aldershot, England: Ashgate.

Klein, G., Kaempf, G., Wolf, S., Thordsen, M., and Miller, T. E. (1997). Applying decision requirements to user-centered design. International Journal of Human-Computer Studies, 46, 1-15.

Klein, G. and Militello, L. G., (2001). Some guidelines for conducting cognitive task analysis. In E. Salas (Ed), *Advances in Human Performance and Cognitive Engineering Research, Volume 1* (pp. 163-199). City, State: Elsevier Science Ltd.

Klein, G., Moon, B., and Hoffman, R. R. (2006a, July/August). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems*, pp. 22-25.

Klein, G., Moon, B., and Hoffman, R.R. (2006b, July/August). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems*, pp. 22-26.

Klein, G., Orasanu, J., Calderwood, R., and Zsambok, C. E. (Eds.) (1993). *Decision making in action: Models and methods*. Norwood, NJ: Ablex Publishing Corporation.

Klein, G., Rasmussen, L., Mei-Hua, L., Hoffman, R.R., and Case, J. (2014). Influencing preferences for different types of causal explanation for complex events. *Human Factors, 56,* 1380-1400.

Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., and Hollnagel, E. (May/June 2003). Macrocognition. *IEEE Intelligent Systems,* pp. 81-85.

Klein, G., Schmitt, J., McCloskey, M., Heaton, J., Klinger, D., and Wolf, S. (1996). A decision-centered study of the regimental command post (Final Contract USC P.O. 681584 for the Naval Command, Control and Ocean Surveillance Center, San Diego, CA). Fairborn, OH: Klein Associates Inc.

Klein, G., Snowden, D. and Chew, L.P. (2007). Anticipatory thinking. *Proceedings of the 8th Conference on Naturalistic Decision Making*, Pacific Grove, CA.

Klein, G., Snowden, D., and Chew, L.P. (2011). Anticipatory thinking. In K. L. Mosier & U.M. Fischer (Eds.) *Informed by knowledge: Expert performance in complex situations* pp. 235-245). New York: Psychology Press.

Klein, G. and Thordsen, M. L. (1989). Recognitional decision making in C^2 organizations. In *Proceedings of the 1989 Symposium on Command-and-Control Research* (pp. 239-244). Science Applications International Corporation, McLean, VA.

Klein, G. and Thordsen, M. L. (1991). Representing cockpit crew decision making. In R. S. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 1026-1031). Columbus, OH: The Ohio State University.

Klein, G., and Weitzenfeld, J. (1982). The use of analogues in comparability analysis. *Applied Ergonomics*, *13*, 99-104.

Klein, G. A. and Woods, D. D. (1993). Conclusions: Decision making in action. In Klein, G. A., Orasanu, J., Calderwood, R., and Zsambok, C. E. (Eds.), *Decision making in action: Models and methods* (pp. 404-411). Norwood, NJ: Ablex.

Klein, G., Woods, D. D., Bradshaw, J. D., Hoffman, R. R., and Feltovich, P. J. (November/December 2004). Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE: Intelligent Systems*, pp. 91-95.

Klein, L. (1994). Sociotechnical/organizational design. In W. Karwowski and G. Salvendy (Eds.), *Organization and management of advanced manufacturing* (pp. 197-222). New York: John Wiley.

Kuipers, B. (1987). New reasoning methods for artificial intelligence in medicine. *International Journal of Man- Machine Studies, 26*, 707-718.

Kuipers, B., and Kassirer, J. P. (1987). Knowledge acquisition by analysis of verbatim protocols. In A. L. Kidd (Ed.), *Knowledge acquisition for expert systems: A practical handbook* (pp. 45-71). New York: Plenum Press.

Kupiers, B., and Kassirer, J. P. (1984). Causal reasoning in medicine: Analysis of a protocol. *Cognitive Science, 8,* 363-385.

Kupiers, B., and Kassirer, J. P. (1983). How to discover a knowledge representation fork causal reasoning by studying an expert physician. In *Proceedings of the 8th International Conference on Artificial Intelligence* (pp. ###-###). Los Altos, CA: Morgan Kaufman.

Kuipers, B., Moskowitz, A. J., and Kassirer, J. P. (1988). Critical decisions under uncertainty. *Cognitive Science*, *12*, 177-210.

Lichtensetin, S., Fischhoff, B., and Phillips, L. D. (1982). Calibration of probabilities: The state of the art in 1980. In D., Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Lipshitz, R. (1989). *Decision making as argument-driven action.* Boston: Boston University Center for Applied Social Science.

Lipshitz, R. (1993). Converging themes in the study of decision making in realistic settings. In G. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok, C. E. (Eds.), *Decision making in action: Models and methods* (pp. 103-137). Norwood, NJ: Ablex.

Lipshitz, R, and Ben Shaul, O. (1997). Schemata and mental models in recognition-primed decision making. In C. E. Zsambok and G. Klein (Eds.), *Naturalistic decision making* (pp. 293-304). Mahwah, NJ; Erlbaum.

Lipshitz, R., Klein, G., Orasanu, J., and Salas, E. (2001). Rejoinder: A Welcome Dialogue--and the Need to Continue. *Journal of Behavioral Decision Making, 14*, 385-389.

Miller, R. B. (1953). "A method for man-machine task analysis." Report WADC-TR-53-137. Wright Air Development Center, Wright Patterson Air Force Base, OH.

Miller, T. E., Wolf, S. P., Thordsen, M. L., and Klein, G. (1992). "A decision-centered approach to storyboarding anti-air warfare interfaces." Report under Contract No. N66001-90-C-6023 for the Naval Command, Control and Ocean Surveillance Center, San Diego, CA. Fairborn, OH: Klein Associates Inc.

Miller, T. E., and Woods, D. D. (1997). Key issues for naturalistic decision making researchers in system design. In C. E. Zsambok and G. Klein (Eds.), *Naturalistic decision making* (pp. 141-150). Mahwah, NJ; Erlbaum.

Militello, L. G., and Hutton, R. J. B. (1998). Applied Cognitive Task Analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics* (Special Issue on Task Analysis) *41,* 1618-1641.

Militello, L.G., and Lim, L. (1995). Patient assessment skills: Assessing early cues of necrotizing enterocolitis. *The Journal of Perinatal and Neonatal Nursing*, *9*, 42-52.

Militello, L.G., Patterson, E.S., Bowman, L., Wears, R. (2007). Information Flow During Crisis Management: Challenges to coordination in the emergency operations Center. *Cognition, Technology, and Work Special Issue on Large-Scale Coordination, 9*, 25-31.

Militello, L.G., Patterson, E.S., Wears, R. Snead. A. (2005) Emergency Operations Center Design to Support Rapid Response and Recovery. *Proceedings of the Working Together: RandD Partnerships in Homeland Security Conference*. Washington DC: Department of Homeland Security

Militello, L.G., Quill, L., Patterson, E.S., Wears, R., Ritter, J.A. (2005). Large-scale coordination in emergency response. *Proceedings of the 49th Annual Human Factors and Ergonomics Society Meeting.* Santa Monica, CA: Human Factors and Ergonomics Society.

Montgomery, H., Lipshitz, R., and Brehmer, B. (Eds.) (2005). *How professional make decisions.* Mahwah, NJ: Erlbaum.

Moon, B. (2002, May). "Naturalistic Decision Making: Establishing a Naturalistic Perspective in Judgment and Decision-Making Research." Presented at the 19th Qualitative Analysis Conference, McMaster University, Hamilton, Ontario, Canada.

Moore, T. V. (1939). *Cognitive psychology*. New York: Lippincott.

Mosier, K. L. (1997). Myths of expert decision making and automated decision aids. In C. E. Zsambok and G. Klein (Eds.), *Naturalistic decision making* (pp. 319-330). Mahwah, NJ; Erlbaum.

Neville, K., Hoffman, R.R., Linde, C., Elm, W.C. and Fowlkes, J. (2008, January/February). The procurement woes revisited. *IEEE Intelligent Systems*, pp. 72-75.

Mosier, K.L., and Fischer, U. M. (2010). *Informed by Knowledge: Expert Performance in Complex Situations.* Boca Raton, FL. Psychologist Press.

O'Dell, C., and Grayson, C. J. (1998). *If we only knew what we know: the transfer of internal knowledge and best practice*. NY: The Free Press.

Orasanu, J. (1990). "Shared mental models and crew decision making." Report No. 46, Cognitive Sciences Laboratory, Princeton University, Princeton NJ.

Orasanu, J., and Connolly, T. (1993). The reinvention of decision making. In Klein, G. A., Orasanu, J., Calderwood, R., and Zsambok, C. E. (Eds.), *Decision making in action: Models and methods* (pp. 3-20). Norwood, NJ: Ablex.

Orasanu, J., and Fischer, U. (1997). Finding Decisions in natural environments. In C. E. Zsambok, and G. Klein (Eds.), *Naturalistic decision making* (pp. 434-458). Hillsdale, NJ: Erlbaum.

Orasanu, J., and Salas, E. (1993). Team decision making in complex environments. In G. A. Klein, and J. Oarsaun (Eds.), *Decision making in action: models and methods.* Westport, CT: Ablex Publishing.

Orr, J., (1996). *Talking about machines: An ethnography of a modern job.* Ithaca, NY: Cornell University Press.

Payne, J. W., Bettman, J. R., and Johnson, E. J. (1988). Adaptive strategy selection in decision making. *J*

Pennington, N. (1981). "Causal reasoning and decision making: two illustrative analyses." Paper presented at the second annual meeting of the Cognitive Science Society. New Haven, CT.

Pennington, N., and Hastie, R. (1981). Juror decision making models: the generalization gap. *Psychological Bulletin, 89*, 246-287.

Pennington, N., and Hastie, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology; Learning, Memory, and Cognition, 14*, 521-533.

Pennington, N., and Hastie, R. (1993). A theory of explanation-based decision-making. In Klein, G. A., Orasanu, J., Calderwood, R., and Zsambok, C. E. (Eds.), *Decision making in action: Models and methods* (pp. 188-204). Norwood, NJ: Ablex.

Pillsbury, W. B. (1929). *The history of psychology.* New York: W. W. Norton and Co.

Pitz, G. F., and Sachs, N. J. (1984). Judgment and decision: theory and application. The *Annual Review of Psychology, 35*, 139-163.

Pliske, R., Crandall, B., and Klein, G. (2004). Competence in weather forecasting. In K. Smith, J. Shanteau, and P. Johnson (Eds.), *Psychological investigations of competent decision making* (pp. 40-70). Cambridge: Cambridge University Press.

Potter, S. S., Roth, E. M., Woods, D. D., and Elm, W. C. (2000). Bootstrapping multiple converging cognitive task analysis techniques for system design. In J. M. Schraagen and S.F. Chipman (Eds.), *Cognitive task analysis* (pp. 317-340). Mahwah, NJ: Erlbaum.

Raffia, H. (1986). *Decision analysis: Introductory lectures on choices under uncertainty.* Reading, MA: Addison-Wesley.

Rasmussen, J. (1979). *On the structure of knowledge: A morphology of mental models in a man-machine context.* (Report No. RIS∅-M-219). Roskilde, Denmark: Riso National Laboratory.

Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics, 13,* 257-266.

Rasmussen, J. (1992). The use of field studies for design workstations for integrated manufacturing systems. In M. G. Helander and N. Nagamachi (Eds.), *Design for manufacturability: A systems approach to concurrent engineering and ergonomics* (pp. 317-338). London: Taylor and Francis.

Rasmussen, L.J., Sieck, W.R., & Hoffman, R.R. (in press). Cultural knowledge for intelligence analysts: Expertise in cultural sensemaking. *American Intelligence Journal, 31,* 28-37.

Ribot, T. (1890). *Psychology of attention.* Chicago: Open Court.

Ross, K., and Shafer, J. (2006). Naturalistic decision making and the study of expertise. In A. Ericsson, N. Charness, P. Feltovich, and R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance*. Cambridge: Cambridge University Press (in production).

Roth, E. M. (1997). Analysis of decision making in nuclear poweer plant emergencies: an investigation of aided decision making. In C.E. Zsambok and G. Klein (Eds.), *Naturalistic decision making* (pp. 175-182). Mahwah, NJ; Erlbaum.

Salas, E., Cannon-Bowers, J. A., and Johnson, J. H. (1997). Ho can you turn a team of experts into an expert team?: Emergency training strategies. In C. E. Zsambok and G. Klein (Eds.), *Naturalistic decision making* (pp. 359-370). Mahwah, NJ; Erlbaum.

Salas, E., and Fiore, S. (2004). *Team cognition.* Washington, DC: APA Books.

Salas, E., and Klein, G. (2001). *Linking expertise and naturalistic decision making.* Mahwah, NJ: Lawrence Erlbaum Associates.

Salas, E., Rosen, M., Burke, S., Goodwin, G. F., and Fiore, S. M. (2006). The making of a dream team: When expert teams do best. In K. A., Ericsson, N. Charness, P. Feltovich, and R. Hoffman, (Eds.). *Cambridge handbook of expertise and expert performance* (pp. 439-453). New York: Cambridge University Press.

Salas, E., Stagl, K.C., and Burke, C.S. (2004). 25 Years of team effectiveness in organizations: Research themes and emerging needs. *International Review of Industrial and Organizational Psychology, 19*, 47-91.

Salas, E., Stagl, K.C., Burke, C.S., and Goodwin, G.F. (in press). Fostering team effectiveness in organizations: Toward an integrative theoretical framework of team performance. In W. Spaulding, and J. Flowers (Eds.), *Modeling complex systems: Motivation, cognition and social processes.* Lincoln, NE: University of Nebraska Press.

Sanders, M. S., and McCormick, E. J. (1992). *Human factors in engineering and design.* New York: McGraw Hill.

Schaafstal, A., Schraagen, J. M., and van Berlo, M. (2000). Cognitive task analysis and innovation of training: The case of structured trouble shooting. *Human Factors, 42,* 75-86.

Schraagen, J. M. C. (2006). Task analysis. In K. A. Ericsson, N. Charness, P. J. Feltovich, and R. R. Hoffman (Eds.) *Cambridge handbook of expertise and expert performance* (pp. 185-202). New York: Cambridge University Press.

Schraagen, J. M. C., Chipman, S. F., and Shalin, V. L. (Eds.) (2000). *Cognitive task analysis*. Hillsdale, NJ: Erlbaum.

Schraagen, J.M., Klein, G. and Hoffman, R. (2008). The macrocognition framework of naturalistic decision making. In J.M. Schraagen, L.G. Militello, T. Ormerod and R. Lipshitz (Eds.), *Naturalistic decision making and macrocognition* (pp. 3-25). Aldershot, England: Ashgate.

Schraagen, J. M. C., Militello, L. G., Ormerod, T., and Lipshitz R. (Eds). (2007). *Naturalistic decision making and macrocognition*. Aldershot, UK: Ashgate Publishing Limited.

Serfaty, D., MacMillan, J., Entin, E. B., and Entin, E. E. (1997). The decision making expertise of battle commanders. In G. A. Klein and C. E. Zsambok (Eds.) *Naturalistic decision making* (pp. 2433-246). Hillsdale, NJ: Lawrence Erlbaum Associates.

Sethi-Iyengar, S., Huberman, G., and Jiang, W. (2004). How much choice is too much? Contributions to 401(k) retirement plans. In Mitchell, O. S. & Utkus, S. (Eds.) *Pension Design and Structure: New Lessons from Behavioral Finance*. Oxford: Oxford University Press, 83-95.

Shanteau, J. (1984). Some unasked questions about the psychology of expert decision makers. In M. E. El Hawary (Ed.), *Proceedings of the 1984 IEEE Conference on Systems, Man, and Cybernetics* (pp. 23-45). NY: IEEE.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes, 53*, 252-266.

Shanteau, J., and Stewart, T. R. (1992). Why study expert decision making?: Some historical perspectives and comments. *Organizational Behavior and Human Decision Processes, 53*, 95-106.

Stanard, R., Uehara, M., and Hutton, R. J. (2003). "Year One Final Report: Decision-Centered Design: Principles and Processes." Report to the Advanced Decision Architectures Collaborative Alliance, Army Research Laboratory, Aberdeen MD.

Stewart, T. R. (2001). Improving reliability of judgmental forecasts. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 81-106): Kluwer Academic Publishers.

Taynor, J., Crandall, B., and Wiggins, S. (1987). "The reliability of the critical decision method." Technical Report on Contract No. MDA903-86-C-0170, U.S. Army Research Institute, Alexandria, VA. Klein Associates, Inc., Yellow Springs, OH.

Taynor, J., Klein, G., and Thorsdsen, M. (1987). Distributed decision making in wetland firefighting. Report, Contract MDA-903-85-C-0327, U. S. Army Research Institute, Alexandria, VA.

Tyzska, T. (1985). simple decision strategies versus multi-attribute utility theory approach to complex decision problems. *Praxiology Yearbook, 2*, 159-172.

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work.* Mahwah, NJ: Erlbaum.

Warwick, W. and Hutton, R. (2007). Computational and theoretical perspectives on recognition primed decision making. In R. R. Hoffman (Ed.), *Expertise out of context: Proceedings of the Sixth International Conference on Naturalistic Decision Making* (pp. 429-451). Boca Raton, FL CRC Press.

Weitzenfeld, J. S., and Klein, G. (1979). "Analogical reasoning as a discovery logic." Technical Report on Contract No. F49620-79-C-0179, Air Force Office of Scientific Research, Bolling AFB, DC. Klein Associates, Inc., Yellow Springs, OH.

Weitzenfeld, J. S., Riedl, T. R., Freeman, J. T., Klein, G., and Musa, J. (1991). Knowledge elicitation for software engineering expertise, In J. E. Tomayko (Ed.), *Proceedings of the 1991 Software Engineering Education (SEI) Conference* (pp. 283-296). New York: Springer-Verlag.

Woods, D. D. (1993). Process tracing methods for the study of cognition outside of the experimental psychology laboratory. In G. A. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok (Eds.), *Decision making in action: Models and methods,* (pp. 228-251). Norwood, NJ: Ablex.

Woods, D. D. (2000). *Lessons from beyond human error: Designing for resilience in the face of change and surprise.* NASA Design for Safety Workshop, October 8-10, NASA-Ames Research Center, Moffett Field, CA.

Woods, D. D. (2002). Steering the reverberations of technology change on fields of practice: Laws that govern cognitive work. In *Proceedings of the 24th Annual Meeting of the Cognitive Science Society* (pp. 1-14). Austin, TX: Cognitive Science Society.

Woods, D.D. (2015). Four concepts for resilience and the implications for the future of resilience engineering. *Reliability Engineering and System Safety*, in press.

Woods, D. D. and Branlat, M. (2011). Basic patterns in how adaptive systems fail. In E. Hollnagel, J. Pariès, D. D. Woods, and J. Wreathall (Eds.), *Resilience Engineering in Practice* (pp.127-144). Farnham, UK: Ashgate.

Woods, D. D. and Hollnagel, E. (2006). *Joint cognitive systems: Patterns in cognitive systems engineering.* Boca Raton FL: Taylor & Francis.

Woods, D.D. and Patterson, E.S. (2000). How Unexpected Events Produce an Escalation of Cognitive and Coordinative Demands. In P. A. Hancock and P. Desmond (Eds.), *Stress workload and fatigue* (pp. 290-302). Mahwah, NJ: Lawrence Erlbaum Associates.

Woodson, W. E., Tilman, B., and Tilman, P. (1992). *Human factors design handbook* (2nd ed.). New York: McGraw-Hill.

Zachary, W., Hoffman, R.R., Crandall, B., Miller, T. and Nemeth, C. (2012, March/April). "Rapidized" cognitive task analysis. *IEEE: Intelligent Systems,* pp. 61-66.

Zakay, D., and Wooler, S. (1984). Time pressure, training, and decision effectiveness. *Ergonomics, 27*, 273-284.

Zsambok, C. E. (1997). Naturalistic decision making research and improving team decision making. In C. E. Zsambok and G. Klein (Eds.), *Naturalistic decision making* (pp. 111-120). Mahwah, NJ; Erlbaum.

Zsambok, C. E., Beach, L. R., and Klein, G. (1992). A literature review of analytical and naturalistic decision making. Fairborn, OH: Klein Associates Inc.

Zsambok, C. E., Kaempf, G. L., Crandall, B. and Kyne, M. (1996). *A cognitive model of a prototype training program for OJT Providers.* Technical Report (Contract no. MDA903-93-C-0092, U. S. Army Research Institute, Alexandria, VA). Klein Associates, Inc., Fairborn, OH.

Zsambok, C. E., and Klein, G. (1997). *Naturalistic decision making.* Mahwah, NJ: Lawrence Erlbaum Associates.

Appendix A

Background and History on "Naturalistic Decision Making

Summary

Naturalistic decision making refers to the study of decision making in domains characterized by time stress, high stakes, vague goals, uncertainty, multiple players, organizational constraints, and dynamic settings. This approach to the study of decision making has had considerable applied impact as well as important theoretical contribution. In the 1980s, studies of firefighters and neonatal intensive care nurses using retrospective interviews to explore actions, goals, and plans, as well critical cues that influenced decision making were conducted. The outcomes of these studies had an impact on the firefighting and nursing communities that participated. Investigators succeeded in aiding interviewees in articulating cues and cue patterns that had not previously been documented—and were subsequently integrated into training programs.

On a broader scale, these early studies also provided important evident that the decision-analytic model has important limitations. The decision-analytic model had been developed and extensively applied in economics and business decision making within the Judgment and Decision Making (JDM) paradigm. The JDM paradigm focuses on issues such as optimal strategies prescribed by probability theory or expected utility theory, accuracy (or lack thereof) of judgments, reasoning biases and limitations in the human ability to evaluate the probabilities of events. The decision-analytic model was widely prescribed as being the best method for making decisions.

In the 1980s some researchers began to react against the JDM paradigm. Researchers reacted against the characterization of human cognition as flawed. The prescriptive nature of the decision-analytic model became a target for criticism. Studies using CTA-CFR methods suggested that decision making in the real world could not be reduced to a single moment of choice after all the facts had been gathered, but was constructed through an incremental process. CTA-CFR studies suggest a series of decisions-like points along a timeline in which actions are taken and options exist, but no concurrent evaluation of options occurs.

A number of limitations of the JDM approach were articulated. Time critical domains do not allow formal generation and evaluation procedures. Experts rarely reported considering more than one option at a time. Further, research has shown that people are not very good at either generating lists of options or at systematically evaluating options. Another debate focused on the use of "hit rate" to assess proficiency. Studying hit rates allows for the use of statistical methods to for diagnostic evaluation in of decision-making skill. NDM researchers, however, take the perspective that linear modeling and focusing solely on hit rates ignores all the richness of proficient knowledge and skill.

The NDM paradigm has come to be defined by three distinguishing characteristics: 1) focus on examination of decision making in everyday situations, both routine and non-routine situations, both simple and complex, 2) a focus on decision making by experienced, knowledgeable individuals, and 3) the examination of decision-making in "real world" job contexts. Given these foci, most NDM researchers investigate "ill-structured" problems and

domains, uncertain and dynamic environments, situations which involve goal conflict, scenarios involving time pressure and high risk, and team or group problem-solving.

CTA methods commonly used by the NDM community include the Critical Decision Method (CDM), the Knowledge Audit, and Goal-Directed Task Analysis. Of these, the CDM has been applied and explored most thoroughly. CDM was initially developed during studies of decision making in fireground command. Researchers adapted Flannagan's (1954) critical incident technique to focus on critical decisions. In a CDM interview, the expert is asked to recall an critical incident, and the interviewer walks through the incident several times with the interviewee, unpacking more of the story and more detail with each sweep. The method has been refined over time and explored for validity from a range of perspectives.

The Knowledge Audit was developed to complement incident-based techniques such as the CDM. Based on an understanding of human expertise, question probes were developed to obtain examples of various aspects of expertise as they are instantiated in a specific domain. Rather than eliciting one critical incident that is thoroughly explored as in the CDM, the knowledge audit elicits a series of incidents, illustrating aspects of expertise such as diagnosing and predicting, situation awareness, improvising, metacognition, recognizing anomalies, and compensating for technology limitations. Often the Knowledge Audit is used early in a study to obtain an overview of the knowledge and skills needed in a specific domain. The Knowledge Audit has also been used to explore differing levels of proficiency.

Goal-Directed Task Analysis (GDTA) is another interview technique used in NDM research. GDTA interviews are organized around the goals the decision maker must achieve, and the information needed to achieve those goals. GDTA does not restrict task description to a linear, sequential series of activities or even hierarchies. GDTA takes into account the characteristics of complex cognitive systems including conflicting goals and the processing on information in ongoing situations. This method is similar to Hierarchical Task Analysis in that goals and subgoals are elicited and represented in a graphical decomposition.

Two main theoretical contributions of the NDM community include the Recognition-Primed Decision-Making (RPD) model and integrated theory of Situational Awareness. The RPD model was based on interview data obtained from experienced firefighters, and later refined expanded based on interview data collected from experienced critical care nurses and experts in a range of other domains. RPD was articulated in reaction against the analytic decision making. Rather than generative a range of options and comparing them to select the best one at a specific decision point, firefighters reported that they rarely had time to consider alternative options. Instead, they seemed to rely on matching the current situation to a typical course of action based on internal prototypes or analogs. In the RPD model, the decision maker spends most of the time available assessing the situation rather than evaluating options.

In short, the expert recognizes the situation (generally as an analog or prototype). Byproducts of the recognition include relevant cues, expectancies, plausible goals, and a typical course of action, all of which become activated based on the recognition of the situation. In the simplest form of RPD, the typical course of action is implemented without conscious deliberation. Variations include situations in which the decision maker does not immediately recognize the situation as familiar. In this case, the experienced decision maker is likely to engage in feature matching or storybuilding to assess the situation. After assessing the situation, the same four byproducts, including a typical course of action become evident and the course of action is implemented. In a third variation, if time is available, the expert may pause before implementing the course of action to mentally simulate or imagine how events will unfold. As a result of this mental simulation, the course of action may be refined or rejected and another selected. However, even in this third variation of RPD, option are not compared, instead they are considered serially until an acceptable course of action is generated.

A second theoretical contribution of NDM has been in moving from the notion of attention to an integrated theory of reasoning termed Situation Awareness (SA). The theory of SA emphasizes that attention involves not just the detection of isolated signals, stimuli or cues, or even the perception of static objects, but the on-going awareness of one's environment. Three levels of SA have been posited: 1) Meaningful interpretation of data, resulting in information; 2) comprehension of information, resulting in a mental model, or higher order understanding prioritized according to how it related to achieving goals; and 3) the mental or imaginal projection of events into a possible future.

Both of these theories have led to approaches to the design of information technologies. Situation-Awareness Oriented Design relies on Goal-Directed Task Analysis (GDTA) and the theory of SA to form an approach to designed technologies intended to support active organization of information, active search for information, active exploration of information, reflection on the meaning of information, and evaluation and choice among action alternatives. In this approach, SA requirements analysis is conducted using GDTA. Next, design principles are used to translate SA requirements into ideas for system design. In the final step, the design is tested using the Situation Awareness Global Assessment Technique (SAGAT).

Decision-Centered Design is an approach motivated by the RPD perspective, and is intended to focus the development of technologies on supporting decision making. The DCD process begins with the identification of individuals who will be users of the new technology. Ideally, these are experts in the domain at hand, and analysts are able to obtain a rich understanding of their needs and requirements. The analysis portion of DCD involves revealing and studying the challenging and critical aspects of jobs. There is a working assumption that 80% of the problems can be solved by understanding and improving the toughest 20% of the cognitive work.

Finally, it is important to mention the emerging notion of cognition, an idea introduced to capture the phenomena of decision making that occur in natural settings as opposed to artificial laboratory settings. The notion of macrocognition has dovetailed with NDM research in the search for an integrated model of reasoning. Macrognition refers to the perspective that in a real-world context, it makes sense to refer to processes such as problem detection, sensemaking, re-planning, and mental simulation, which are continuous and interacting. This is in contrast to microcognition, which attempts to reduce mental operations to hypothetical building blocks (i.e., attentional switching, sensation, memory contact, recognition) placed into causal strings. The microcognitive approach is perhaps most appropriate for probing cognition at the millisecond level of causation, rather than in the larger context of on-the-job performance. The study of micro- and macrocognition are complementary. Lab-based studies of microcognition is needed in parallel with the study of emergent macrocognitive phenomena typically studied in field settings. Both micro and macrocognitive research findings have implications for the design of technologies.

Origins of This Community of Practice

Discussions of the origins of the NDM paradigm appear in Klein, et al. (1993), Moon (2002), and Ross and Shafer (2006). NDM as a community of practice has no formal society, but is sustained by meetings and common interests. It began with the first conference in 1989 in Dayton Ohio, at which group of researchers who were studying different domains for different reasons found a common and seemingly distinctive set of goals and methods. At that meeting, Judith Orasanu, a leading human factors psychologist at NASA, laid out the key features of the NDM attempt to "redefine decision making" (Orasanu and Connolly, 1993) through the study of "real world" decision making by domain experts working at challenging tasks that that are dynamic, ill-structured, and high-stakes. The 1989 meeting was intended as a workshop to allow sharing of recent results and interests, but it sparked demand for follow-on gatherings. The NDM community has met every 2-3 years since then, alternating between North American and European venues. Each of the NDM meetings has generated a book describing the research and the ideas of the conference participants (Hoffman, 2007; Klein et al., 1993; Zsambok and Klein, 1997; Flin, Salas, Strub and Martin, 1997; Salas and Klein, 2001; Montgomery, Lipshitz, and Brehmer, 2005; Schraagen, 2007; Militello, Ormerod, and Lipshitz, 2007; Schraagen, 2008). Many NDM researchers gather every year as part of the Cognitive Ergonomics and Decision Making Technical Group within the Human Factors and Ergonomics Society, and at meetings on Situation Awareness.

What triggered this?

A Study of Firefighters

In the mid-1980s, the US Army Research Institute funded a research project to study decision making in time-pressured, high-risk settings. This led to a series of studies in which interviews were conducted with professional urban and forest firefighters (Calderwood, Crandall, and Klein, 1987; Klein, Calderwood, and Clinton-Cirroco, 1986). In these retrospective interviews, the participant recounted previously-experienced cases that were rare or that involved tough, challenging decisions. Participants in these studies included individuals who had about a decade of experience (i.e., rank of captain or above), and individuals who had only one or two years of experience as firefighters and no experience as fireground commanders (i.e., they were newly-promoted lieutenants). In the knowledge elicitation task, the participants recalled cases from their past experience, described he events in terms of timelines, and answered probe questions about each decision point on the timeline (e.g., "What information did you need at that point?," "What were you seeing at that point?," "What were your options at that point?").

The results included information about the experts' actions, goals, and plans. The probe questioning yielded information about the cues to which the experts attend, and information about how the cues were linked to causal relations, actions, and plans. Investigators were able to specify many of the important cues in various types of firefighting situations—something that had not previously been done to such an extent. Some of the cues and cue patterns that were revealed were ones that the expert has never explicitly deliberated or specified. For example, in the initial description of one of his experiences, a firefighter initially explained that he had a "sixth sense" for judging the safety of a fire ground (i.e., a burning roof). Upon the subsequent sweep through the retrospective recall, using the probe questions, the expert "discovered" the perceptual pattern that he relied upon, involving such things as smoke color and the feel of a "spongy" roof. Another findings was that the experts did not spent much time generating and evaluating options. Indeed, in this high-pressure decision making situation, the deliberation of options is not an option: There's no time. Yet, the experts were able to make good decisions, many of them at scales including small scale (e.g., where is the seat of the fire?) and larger scale (e.g., when to call in extra tanker trucks).

A Study of Neonatal Intensive Care Nurses

The experience of nursing instructors had been that proficient nursing skill and knowledge is difficult for the expert to access and articulate, and operates tacitly in the course of decision making. In a study conducted by Beth Crandall and her colleagues Crandall and Calderwood, 1989; Crandall and Gamblian, 1991), a group of 17 expert nurses performed detailed situation assessments for 24 cases of neonatal sepsis. From their accounts, Crandall, et al. generated a description of assessment procedures and a list of indicators (perceptual cues and information from telemetry) of the physiological changes that occur in neonates over the course of sepsis. Cues included color change (pale tone, grey tone, paleness in extremities), apnea or brachycardia (frequency of episodes increases over time), and lethargy (patient is sleepy or listless, limp muscle tone, unresponsiveness). Presumably, all of these important cues had already been spelled out and thoroughly analyzed in the medical textbooks used in clinical training. To test this hypothesis, the three leading texts and manuals, and some of the associated literature in periodicals, were examined for their descriptions of neonatal sepsis in terms of its critical indicators. The finding was that many of the critical indicators discussed in the medical literature had not been mentioned at all by the expert nurses during the situation assessment knowledge elicitation task (e.g., elevated temperature, vomiting, seizures, jaundice).

Furthermore, some of the indicators that were important to the expert nurses were not mentioned in the medical literature (e.g., muscle tone, "sick" eyes, edema, clotting problems). Many of the discrepancies hinged on the clinical nurse's ability to detect early signs of sepsis, that were manifested as cue configurations rather than individual, salient cues. The medical literature focuses on advanced symptoms. Especially salient was the fact that clinical nurses are especially sensitive to certain symptom co-occurrences (e.g. the co-occurrence of pale skin tone with lethargy and apnea). Also, many of the critical cues upon which the clinical nurses relied involved perceptual judgments and alertness to shifts in the patient's state: "... a nurse would describe a growing concern as the infant became increasingly limp and unresponsive and as the infant's color changed from pink to pale to dingy grey over the course of the shift" (Crandall and Getchell-Reiter, 1993, pp.47-48).

Crandall and Klein (1987b) obtained comparable findings in a study of paramedical treatment of heart attacks. Cue configurations that paramedics rely upon in diagnosing heart attacks prior to the onset of the standard symptoms involve skin features (a blue-grey tone, or loss of pinkness, especially at the extremities; a cold, clammy feel), eye response (glazed, unfocused, dilated), breathing changes (rapid, shallow breathing), and changes in mental state (a confused or anxious mental state). In yet a third study modeled after Crandall's initial research, Militello and Lim (1995) identified individual cues and clusters of cues experienced neonatal intensive care nurses rely on to assess an infants risk for necrotizing enterocolitis. In this case, experienced nurses had learned to watch for indicators of gastro-intestinal distress (i.e., increased girth, aspirates) coupled with early signs of infection (i.e., poor perfusion, change in activity level, temperature instability).

These and other findings set the stage for the initial motivation for the NDM paradigm—a reaction against a paradigm, or community of practice, called "Judgment and Decision-Making" (JDM).

The "Normative" View of Judgment and Decision Making

JDM, a field with origins in the 1960s, focused on such domains as economics and business decision-making, and was concerned with discovering whether humans make decisions in accord with a logical standard for reasoning, such as the optimal strategies prescribed by probability theory or expected utility theory (Edwards, 1965a). A second line of research focused on the accuracy (or lack thereof) of judgments, including judgments made by experts (e.g., Hammond, McClelland and Mumpower, 1980; Hoffman, 1960; Slovic, 1966). A second line of research focused on the flip side: Reasoning biases and limitations in the human ability to evaluate the probabilities of events (Kahneman, Slovic and Tversky 1982; Kahneman and Tversky, 2000). One of the more consistent findings was that simple linear models equal or outperform humans, even experts with years of experience, on a wide variety of judgment tasks (Dawes, 1979; Dawes and Corrigan, 1974; Grove and Meehl 1996; Swets, Dawes, and Monahan, 2000). If one takes the same information that the human has, apply an appropriate weight to each item of information, and add them up, we get a result that is almost guaranteed to be as accurate as the human. If one provides the human with more information, the human does not necessarily get better, and in fact can become more unreliable and less consistent (Stewart, 2001)

The Decision-Analytic Model

According to the Decision-Analytic Model, the good decision maker:

1. Specifies all the objectives, or the criteria for a solution→
2. Lays out all of the alternative actions→
3. Weighs the benefits versus the costs or risks of each alternative ("utility analysis")→
4. Conducts a multiattribute evaluation of the alternatives→
5. Orders the alternatives in terms of their satisfaction of the criteria→
6. Selects one option for implementation→
7. Engages in contingency planning.

This model, or some variation of it, was widely prescribed as being the best method for conducting the decision-making process (e.g., Janis and Mann, 1977; Raffia, 1986). The general decision-analytic model is portrayed in the Figure A.1.

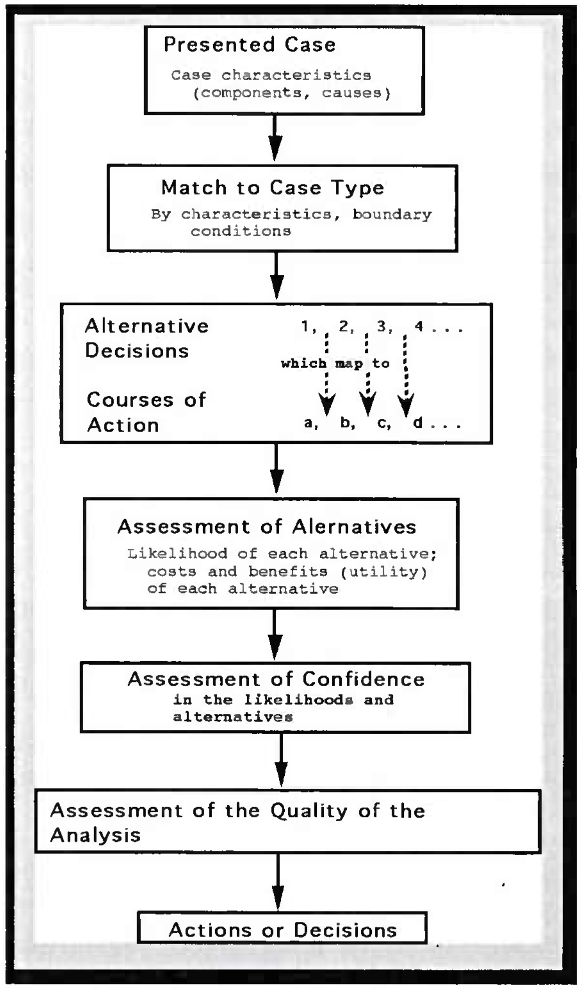


Figure A.1. A depiction of the decision-analytic model of decision making.

An example of decision analysis is a study by Kuipers, Moskowitz, and Kassirer (1988). Their study also illustrates one of the limitations of the decision-analytic model. The domain under investigation was medical diagnosis. The medical records of the test case they utilized permitted a detailed traditional decision analysis, including the construction of a decision tree involving choice points, such as "perform lung biopsy," and the likelihoods of all of the possible scenarios based on the available data (e.g., the likelihood of the patient surviving a biopsy if a fungal infection were present). From the likelihoods (calculated on the basis of the medical literature), the utilities of alternative courses of action could be determined. The decision analysis was compared to results from think aloud problem-solving protocols in which three experts analyzed the case.

The results showed that none of the physicians explicitly considered a particular alternative. Probe questioning revealed that the alternative in question would not have been considered because it was not clinically appropriate. Furthermore, the experts' reasoning never really involved a sequence of laying out all the alternatives and then assessing the likelihoods or calculating the utilities. Rather, the experts "made an initial decision at an abstract level, and then went onto specify it more precisely" (Kuipers, et al., 1988, p. 193). In the terminology of decision trees, they moved from the root to a main branch and only considered more specific alternatives as they proceeded along a particular path or course of action.

According to Janis and Mann (1977, p.11), the failure to engage in a full formal decision analytic process represents a "defect" in decision-making (e.g., the failure to engage in this process as a consequence of time pressure will result in the ineffective use of all of the available information). However, a number of NDM researchers found the attempt to deliberately induce a decision-analytic procedure could actually interfere with decision-making. Lipshitz (1987), for example, analyzed the decision-making protocols of military commanders in terms of the decision analytic model and found that forcing them to engage in a decision analysis distorted their usual strategies and reasoning sequences, and failed to capture the recognitional aspects of command decision-making.

The Rallying Point

In the 1980s, some researchers began to react against the JDM line of work, in part because the work was perceived as expressing a fairly negative view of human cognition.

- People tend to seek evidence that confirms hypotheses and do not look for disconfirming evidence.
- People are unreliable and inconsistent.
- People are said to only consider one or two hypotheses at a time.
- People's judgments of event likelihood are biased by recency, salience, and other factors.

The decision-analytic model, especially its prescriptive nature, became a target for criticisms: "Decisions are not made after gathering all the facts, but rather constructed through an incremental process of planning by successive refinement" (Beach and Lipshitz, 1993, p.21). Furthermore, the decision-analytic model focuses on a final point in the problem-solving process, the moment of choice at which "a decision" is made (cf. Berkeley and Humphreys, 1982). Yet, a new wave of research, using what we would now say was CTA-CFR methods, suggested that problem-solving in real world (versus academic laboratory) contexts involves a number of decision-like points along a timeline, points at which actions are taken and where options existed but there is no concurrent evaluation of options, no single "choice" (cf. Hoffman and Yates, 2005; Isenberg, 1984; Lipshitz, 1989). The only thing that "...makes [an action] a 'decision' is that meaningful options do exist and that the decision maker can articulate them if necessary" (Klein, 1989, p.66).

Illustration of "Naturalistic" Decision Making

A study that illustrates how real-world decision-making can rely on a strategy quite different from the decision-analytic strategy is a project conducted by Nancy Pennington and her colleagues (1981; Pennington and Hastie, 1981, 1988, 1993) on juror decision-making. During the course of a trial, jurors are confronted with a great deal of information, information that can be rich in its implications. Pennington and her colleagues investigated what happens in the reasoning of jurors during trials, and found that a majority of their time is spent creating story-like explanations of the causal events, integrating the evidence and the perceived intentions of the participants in the events—and this the jurors *must* do since evidence is presented in a piecemeal, scrambled sequence. Sometimes jurors will conceive more than one possible story or schema, but they usually settle on one as being more coherent (complete, consistent, plausible). One of the consequences of this reasoning strategy could be tested empirically by engaging research participants (college students who participated in mock trials) in a recognition task involving statements that might, or might not, have actually been presented in evidence. Jurors were more likely to falsely recognize statements if they fit their own constructed story schemas (Pennington and Hastie, 1988).

A judge's final instructions to a jury present jurors with a set of mutually-exclusive alternative decisions, and the task for the jurors is to attempt to match their story-explanations to one of the permitted alternative decisions. Another implication of the juror decision strategy could be tested by comparing the verdict of a juror to that juror's story schema. Jurors who chose different verdicts had constructed different stories and there was a distinct causal configuration to each story structure that mapped to each of the verdict categories. In another experiment, mock trials were composed such that witnesses would present evidence in such a way as to describe the unfolding of the events—the evidence could be presented as a coherent story rather than piecemeal across witnesses or evidence categories. Participants were more likely to render a verdict of guilty if the prosecution's case had been presented in story order, and were the least likely to render a guilty verdict if the defense evidence had been presented in story order. The model that Pennington adduced—which she called "explanation-based decision-making"—places greatest weight on the process of reasoning about evidence as it is gathered, rather than on the evaluation of evidence in a decision-analytic fashion after it has all been gathered.

Additional Rallying Points

The blossoming research on expertise in the 1980-90s, and Klein and Crandall's research on experts seemed to be painting a different picture from that painted by JDM.

Decision Making Under Pressure

The research was suggesting that the JDM approach was incomplete—the deliberation of options was simply not possible in time-critical domains, and in any event did not seem to describe what actually occurs in cognition during time-critical decision-making. Experts rarely reported considering more than one option at a time (Klein, Calderwood, and Clinton-Cirroco, 1986). In domains involving time-pressure it is literally impossible for the decision maker to conduct a formal generation and evaluation procedure. Zakay and Wooler (1984) trained participants in a decision-analytic strategy and found that problem solving could proceed effectively using the strategy if there were no time pressure. But if even moderate time pressure was imposed, the strategy was not beneficial. Over trials, the decision-analysis strategy was truncated and adapted, and eventually replaced by a more "heuristic" strategy (see also Payne, Bettman, and Johnson, 1988).

Decision Making on Unfamiliar Problems

Research suggested an explanation of why people are bad at decision analysis. Many studies had shown that people are not very good at either generating lists of options or at systematically evaluating options (e.g., Gettys, Fisher, and Mehle, 1978; Pitz and Sachs, 1984). Choices and decisions are not systematically based on the notions of utility and optimization (Fischhoff, Goitein, and Shapira, 1982; Simon, 1955). Furthermore, when people are forced to engage in a decision analytic strategy, their evaluations can be subject to strong context effects, and judgments of uncertainty are insensitive to such things as the prior likelihoods of events or outcomes (Fischhoff, Slovic, and Lichtenstein, 1979; Lichtensetin, Fischoff, and Phillips, 1982). The view of NDM was that such results are to be expected when people (mostly, college freshmen) are confronted with artificial problem puzzles, or probability-juggling tasks. The same sort of thing can occur when domain experts are presented with problems that fall outside their domain. In a clever experiment by Tyzska (1985), experienced architects and car designers worked on two problems, one involving choosing an apartment and one involving choosing a car. The architects spent more time using a strategy reminiscent of decision-analysis on the car choice problem, the car designers on the apartment choice problem. When confronted with problems outside their skill set, how could they do anything but a decision-analysis like process, that is, an evaluation of individual courses of action and their costs or benefits (Klein, 1989)?

Missing the Forest for the Trees

Another debate arose concerning the assessment of proficiency in terms of the "hit rate" or correctness of final decisions, an approach common to studies in the paradigm of JDM (Hoffman, et al., 1995). Part of the motivation for the study of hit rates is to support the use of statistical analysis, at least for purposes of diagnostic evaluation of decision-making skill. In linear statistical modeling, which we mentioned earlier in this Chapter, the important features or dimensions of analysis are specified, and their values mapped onto a simple measure of outcome. Analysis of cases reveals weightings for the variables, which serve to specify the regression equation. The predictions of the linear model are then compared with the predictions of domain experts for a set of test cases.

It has been argued that linear modeling ignores all the richness of proficient knowledge and skill. Whether statistical prediction outperforms human prediction can depend critically on the task, the experience level of the decision maker-participants, and the amount and kind of contextual information that is available to the decision maker. A simple linear model, for example, cannot take "broken leg" cues into account. This interesting nomenclature comes from the example of the personality assessment device used to predict whether a John Doe of some particular personality type or lifestyle would go to the movies some time during the next month. The model cannot take into account the consequences of John having just broken a leg. A human (expert or not) could. Whether hit rates are a useful measure for certain purposes (or not), the focus on hit rates, argue NDM researchers, ignores nearly everything else that is important about expertise—perceptual skills, knowledge, and context-sensitivity, and their relation to proficient performance.

The "Inherent Predictability" of Events

Some research had shown that linear regression equations can outperform human expert (Dawes, 1979), even when the expert insists that the problems are complex. However, this finding has generally obtained for domains in which the expert's task is to predict either individual human behavior (e.g., diagnosis in clinical psychology, prediction of recidivism by parole officers), or stochastic aggregates of human behavior (e.g., the stock market, prediction of economic trends), and for tasks involving a lack of feedback, the assessment of dynamic situations, and a lack of decision aids (Shanteau, 1988, 1992). Stewart (2001) argued that the fallibility of judgment, including expert judgment, is linked to the inherent predictability of events:

> Predictability determines the maximum possible accuracy of judgments (either predictions, prognoses, or diagnoses), given currently available information. Predictability can be reduced either by inherent randomness or by inadequate or imprecise information, or both. Clearly, problems differ with regard to predictability. Predictability is important not only because it is a ceiling on the potential accuracy of judgment, but also because it affects the reliability of judgments (Stewart, personal Communication, 2004).

People tend to respond to less predictable environments by behaving less consistently (Brehmer 1978; Camerer 1981; Harvey 1995). From this, Stewart predicted that there will be greater disagreement among expert judgments for less predictable events, and that the performance gap between (perfectly reliable) mathematical models and (less reliable) human judgment will increase as inherent predictability decreases. These predictions were supported in a meta-analysis of studies that compared humans with linear predictive models. Stewart, Roebber, and Bosart (1997) found that for high-predictability tasks (e.g., weather forecasters' predicting precipitation or temperature), the performance of humans nearly equaled that of models, and there was close agreement among experts. For low predictability tasks—pathologists predicting survival time for patients who had died of Hodgkin's disease (Einhorn, 1972) and clinical psychologists judging psychosis or neurosis in patients (Goldberg, 1965)—the performance of the best judges nearly matched a linear model, but there was a greater downward range of performance, so most experts performed worse than linear models.

The NDM Paradigm Defined

NDM came to be defined as a paradigm involving:

1. A focus on the examination of decision-making in "everyday" situations, both routine and non-routine situations, both simple and complex,
2. A focus on decision making by experienced, knowledgeable individuals.
3. The examination of decision-making in "real world" job contexts anchors NDM in the study of decision-making in domains that are especially important to business, government, and society at large.

These features distinguish NDM from traditional academic psychology, not because NDM work *must* take place in "the field" (although it often does); not because NDM work looks *only* at domains of practice that are important to business, government, and society (even though much of it does); not because laboratory research *must* eliminate all real-world complexity (it need not); not because NDM research *always* involves looking at experts (though it often does).

Rather, NDM is distinguished because traditional academic research tends to utilize simplified, artificial context-free problems, artificial tasks that occur only in the laboratory, and college undergraduates who serve, more or less willingly, as "subjects" in cognitive research.

Taken together, these foci serve to outline the interests of most NDM researchers, interests in such topics as "ill-structured" problems and domains, reasoning in uncertain and dynamic environments, reasoning in situations where goals come into conflict, reasoning under stress due to time pressure and high risk, and team or group problem-solving (see for example, Beach, et al., 1997; Christiansen-Szalanski, 1993; Cohen, 1993; Flin, et al., 1997; Hammond, 1993; Klein, 1993; Klein, Orasanu, Calderwood, and Zsambok, 1993; Miller and Woods, 1997; Orasanu and Connolly, 1993; Woods, 1993; Zsambok and Klein, 1997). Hence, reports at the NDM conferences have involved, for example, studies of medical reasoning, of the skills of fighter pilots, of the use of cognitive tasks analysis and other methods to reveal the knowledge and skills of experts, etc. One goal of NDM research is to discover how people actually make real decisions in real situations. The goal is not to mold human decision-making into normative or prescriptive models (such as the decision-analytic model) (Cohen, 1993).

Even the fundamental concept of the "decision" is brought into question (Hoffman and Yates, 2005). It is not regarded as a thing that is "made," as a single point that is somehow especially privileged in the analysis of problem-solving. Rather, problem-solving is described in terms of the dynamic assessment of situations and the incremental refinement of awareness and action plans. This resonates with the work of Jens Rasmussen and his colleagues (Rasmussen, 1993, p.171), which regards decision-making as a continuous control task rather than the resolution of individual conflicts (see Chapter 10).

A goal of NDM research is to generate methods and technologies that would be useful in supporting the effective exercise of expertise and the preservation and dissemination of expertise. We turn now to a discussion of those methods.

Cognitive Task Analysis Methods that Have Emerged from the NDM Paradigm

Klein's early research (Klein, 1982, 1987; Klein and Weitzenfeld, 1982; Weitzenfeld and Klein, 1979) was on analogical problem-solving by avionics engineers. In the "comparability analysis" procedure that the engineers follow, the reliability and maintainability of new aircraft components or systems are predicted on the basis of historical data about functionally or structurally similar components on older aircraft. Klein and Weitzenfeld had expert avionics engineers perform this familiar task for some test cases (e.g., the specifications for the hydraulics system on a new airplane). As the experts conducted a comparability analysis, they were prompted with a set of pre-planned interview questions. The results were clear: In this task, reasoning by analogy was built-in. That is, new cases were solved by comparison to past cases.

Assuming that this style of reasoning would not be unique to avionics engineering, Klein and his colleagues went on to study other domains.

Evaluation and Refinement of the Critical Decision Method

Earlier in this chapter we summarized the seminal studies on fire ground commanding and the key findings of those studies. In those and subsequent studies, the method of structured retrospection was refined and tested further. The method was related to the "critical incident method" that had been used for some time by human factors psychologists and others, especially in the retrospective analysis of accidents (e.g., Flannagan, 1954). Klein et al found that asking for the recall of critical incidents tended to trigger the recall of cases in which lives or property had been lost and did not necessarily involve situations in which expert skill or knowledge had been put to the test. Thus, refinements of the knowledge elicitation method involved focusing on critical *decisions* since it appeared that the recall and analysis of non-routine cases can be a rich source of data about proficient performance (Klein, Calderwood, and MacGregor, 1989, p.465). Hence, Klein et al. dubbed their method the "Critical Decision Method" (CDM).

Unlike in the critical incident procedure, where the recall and the recalled events are relatively close in time, in CDM procedures events will be recalled well after they actually occurred. A study of forest firefighters (Taynor, Klein, and Thorsden, 1987) explored the effects of such delay. An elicitor conducted the interview procedure with a number of experts shortly after each of a number of critical incidents. A subset of the incidents was again assessed in a second interview procedure conducted five months later. A coder who was not present during the initial procedure conducted a detailed content analysis for the second run of the procedure. The resulting reliabilities across experts of the identified timeline decision points averaged at about 82%, with a range of 56% to 100% over elicitors. This finding suggested, as one would expect, that completeness and accuracy of event recall varies from expert to expert over time.

Another validity check involved having more than one coder specify a timeline based on selected transcripts from randomly-selected event recall sessions. For the validity check in the study of urban firefighters, one coder had been the elicitor in the original sessions, and during those sessions he had developed his initial scheme for coding the decision points in the domain. The second coder was unaware of the scheme and had not been present during the initial interviews. The two judges agreed in their identification of between 81 and 100 percent of the decision points in four selected event recall transcripts. Disagreements reflected the tendency of the new coder to identify too many statements as decision points. This finding suggested that the method can be sensitive to the domain knowledge of the elicitor/coder. This too is to be expected, and would obtain for any knowledge elicitation method especially when the data are analyzed by a judge who is relatively unfamiliar with the domain.

The validity check involved not just assessing inter-judge reliability in the identification of decision points, but also reliability in the classification of decisions. Decisions in this domain had been classified into five basic categories, and the same two judges used this category scheme to independently code the decisions. The rate of agreement was about 67%, and although this was above statistical chance, it indicated that coders had difficulty in making unambiguous judgments at this level of detail. Recalculation of "essential agreement" was based on the fact that some categories of decisions were conceptually similar. This yielded an agreement rate of 87 percent.

A similar assessment of the reliability of the classification of decision strategies was conducted for the earlier forest fire fighting study. Again, two independent judges, one of whom had been the elicitor, classified the decision strategies involved in 18 decision points. Overall, for five coding categories the rate of agreement was 74%, with essential agreement being 89 %.

The findings concerning reliability in the classification of decisions suggest, as one would expect, that any highly fine-grained analysis of decisions or strategies will depend to some extent on the ontology preferred by the analyst. Another general conclusion is that experts love to tell stories. Indeed, in some cases practitioners learn on the job by sharing their "war stories," and even report they learn more that way than through their formal instruction (as illustrated by Orr's 1985 study of photocopier technicians). Providing structure and guidance to story-telling permits the interview process to flow more naturally, like a dialogue. Klein, et al. have reported that this is essential in maintaining the expert's cooperation and interest:

> Our goal was to focus the expert on those elements of the incident that most affected decision-making and to structure responses in a way that could be summarized along a specified set of dimensions while still allowing the details to emerge with the [expert's] own perspective and emphasis intact (Klein, Calderwood, and MacGregor, 1989, p.465).

The CDM probe questions are designed to elicit information that is specific and meaningful: strategies and the basis for decisions, and the perceptual cues on which the decision-maker relies—types of information that were not ordinarily the focus in either laboratory research on expertise or applied knowledge elicitation projects.

Although the CDM was created and refined during the era of expert systems and the rising interest in Expertise Studies, the CDM was not intended to be used solely for knowledge elicitation for the study of experts or for the development of expert systems. It was also envisioned as a technique to support training and instructional design, and to support the preservation of corporate experience: "Organizations suffer when they do not properly value their own expertise and when they loose skilled personnel without a chance to retain, share or preserve the knowledge of people who retire or leave" (Klein, et al., 1989, p.471). Indeed, Klein (1992) carried this attitude over to knowledge-based systems, regarding the technology not just as a set of tools for use as decision aids, but also as a tool to support the capture, preservation, and dissemination of the knowledge, skills, and experience of experts. Klein's seminal paper (1992) on "preserving corporate memory" helped usher in a wave of interest in what is now called knowledge management (cf. Brooking, 1999; O'Dell and Grayson, 1998).

The most detailed reviews of of the CDM can be found in Crandall, Klein, and Hoffman (2006), Hoffman, Crandall, and Shadbolt, (1998), Klein (1987, 1993c) and Klein, Calderwood, and MacGregor (1989). Crandall, Klein and Hoffman (2006) provide a detailed protocol for conducting the CDM procedure.

A second empirical method that stemmed from the NDM research is called the "Knowledge Audit."

The Knowledge Audit

This procedure (Klein and Militello, in press; Militello and Hutton, 1998) is based on the psychological research on expertise (see Chi, Feltovich, and Glaser, 1981; Ericsson and Smith, 1991; Hoffman, 1991; Klein and Hoffman, 1993), which has demonstrated the important cognitive factors or knowledge categories that distinguish novices from experts:

- Experts possess an extensive knowledge base, that is conceptually-organized around domain principles and that makes diagnosis and prediction possible.
- Experts are more effective at forming initial mental models of a problem situation, and are more effective at achieving and maintaining a high level of "situation awareness."
- Experts possess better metacognitive skills—they know how to manage information, what inferences to make, how and when to apply principles, how and when improvise,

how to compensate for equipment or display limitations, how to recognize anomalies, and so on.
- Experts are more effective at prioritizing their activities during multitasking situations.

The Knowledge Audit interview attempts to get directly at these aspects of expertise. In other words, the purpose of the Knowledge Audit is to determine what distinguishes experts from non-experts in a particular domain or task within a domain, including: diagnosing and predicting, situation awareness, improvising, metacognition, recognizing anomalies, and compensating for technology limitations. The goal of the Knowledge Audit is not to demonstrate the importance of these factors. Rather, the goal is to identify the specific things that experts in a given domain need to know and skills they need to possess.

The Knowledge Audit procedure is useful as the very first interview in a cognitive engineering project, since it results in a data set that points the researchers to the important domain knowledge and other consisting of incident analyses in which the salient differences between practitioners of differing levels of proficiency (i.e., expert-journeyman-trainee differences). The Knowledge Audit can also be used to study cognitive styles. An example is a study by Pliske, Crandall, and Klein, (2000) in a study of USAF weather forecasters. Like the CDM, Knowledge Audit probe questions focus on the recall of specific, lived experiences. They do not ask for reflection on generic knowledge or skills. Examples for the forecasting study appear in Table A.1.

Table A.1. Probes used in the study of weather forecasting by Pliske, et al. (2000).

Probe	Knowledge/skill of interest
Can you recall and discuss some experiences where part of a situation just "popped" out at you; where you noticed things that others did not catch?	Skill at perceiving cues and patterns
Have there been times when you walked into a situation and knew exactly how things got there and where they were headed?	Skill at situation assessment
Can you recall past experiences in which you: • Found ways of accomplishing more with less • Noticed opportunities to do things better • Relied on experience to avoid being led astray by the equipment	Metacognition skill, the ability to think critically about one's own thinking.

Pliske, et al. interviewed a total of 65 forecasters (of varying degrees of experience). Next, the researchers engaged in a multi-trial sorting task in which they reached a consensus on categories of the reasoning styles they had observed. These categories focused on the forecasters' overall strategic approach to the task of forecasting, their strategy in the use of computer weather models, their process in creating forecasts, their means for coping with data or mental overload, and their metacognition. The categories they identified were dubbed "Scientist," "Proceduralist," "Mechanic," and "Disengaged." Features of the styles are presented in Table A.2.

Table A.2. Features of the four main reasoning styles observed in weather forecasting by Pliske, et al., (2000).

AFFECT	SKILL	ACTIVITIES
SCIENTIST They tend to have had a wide range of experience in the domain, including experience at a variety of scenarios.		
They are often "lovers" of the domain. They like to experience domain events and see patterns develop. They are motivated to improve their understanding of the domain.	They possess a high level of pattern-recognition skill. They possess a high level of skill at mental simulation. They understand domain events as a dynamic system. Their reasoning is analytical and critical. They possess an extensive knowledge base of domain concepts, principles, and reasoning rules. They are likely to act like a Mechanic when stressed, or when problems are easy. They can be slowed down by hard or unusual problems.	They show a high level of flexibility. They spend proportionately more time trying to understand the weather problem of the day, and building and refining a mental model of the weather. They possess skill at using a wide variety of tools. They are most likely to be able to engage in recognition-primed decision-making. They spend relatively little time generating products since this is done so efficiently.
PROCEDURALIST Typically, they are younger and less experienced.		
Some are lovers of the domain. Some like to experience domain events and see patterns develop. Some are motivated to improve their understanding of the domain.	They are less likely to understand domain events as a complex dynamic system. They see their job as having the goal of completing a fixed set of procedures, but these are often reliant on a knowledge base. Their knowledge base of principles of rules tends to be limited to types of events they have worked on in the past.	They spend proportionately less time building a mental model and proportionately more time examining the computer model guidance. They can engage in recognition-primed decision making only some of the time. They are proficient with the tools they have been taught to use.
MECHANIC They sometimes have years of experience.		
They are not interested in knowing more than what it takes to do the job; not highly motivated to improve.	They see their job as having the goal of completing a fixed set of procedures, and these are often not knowledge-based. They possess a limited ability to describe their reasoning. They are likely to be unaware of factors that make problems difficult.	They spend proportionately less time building a mental model and proportionately more time examining the guidance. They cannot engage in recognition-primed decision making. They are skilled at using tools with which they are familiar, but changes in the tools can be disruptive.
DISENGAGED They sometimes have years of experience.		
They do not like their job. They do not like to think about the domain.	They possess a limited knowledge base of domain concepts, principles, and reasoning rules. Knowledge and skill are limited to scenarios they have worked in the past. Their products are of minimally-acceptable quality. They are likely to be unaware of factors that make problems difficult.	They spend most of the time generating routine products or filling out routine forms. They spend almost no time building a mental model and proportionately much more time examining the guidance. They cannot engage in recognition-primed decision making.

These categories were heuristic, intended to inform the creation of decision aids and other technologies, so that they may "fit" each of the styles that were observed. Pliske, et al did not claim that this set is exhaustive, that all practitioners will fall neatly into one or another of the categories, or that similar categories would be appropriate for any other given domain. The analysis of reasoning styles has to be crafted so as to be appropriate to the domain at hand.

Another method of cognitive task analysis that is associated with the NDM community of practice is Goal-Directed Task Analysis (GDTA).

Goal-Directed Task Analysis

GDTA is a form of structured interview that uses probe questions to conduct a top-won analysis of work (see Endsley, 1993, 1995ab; Endsley and Bolte, 2003). GDTA attempts to obtain detailed knowledge of the goals the decision maker must achieve, and the information requirements for working towards those goals. As we will show, GDTA analyses are hierarchical in form, but even though GDTA might be seen as a form of Hierarchical Task Analysis, the historical origins are distinct and the two approaches have differing focal points. HTA begins by stating a goal that a person has to achieve. This is re-described into a set of sub-tasks and a plan (or plans) for conducting the tasks. The unit of analysis for HTA is the sub-task specified by a goal, activated by an input, attained by an action, and terminated by feedback (p. Annett, 2003). In describing each subtask, many attributes of that subtask are laid out, including the goal. Thus, HTA is a goal-relevant analysis of tasks rather than an analysis of goals themselves (see also Kirwan and Ainsworth, 1992). In other words, the two techniques highlight different, but equally important aspects of decision making (Chow, 2007). HTA primarily focuses on actions or behaviors, while GDTA primarily focuses on perceptions (of whether goal states are attained or not). HTA analyses tasks in the context of the task goals, whereas GDTA analyses the goals themselves (Chow, 2007). (There is some circularity here, of course, since many descriptive statements of goals can be regarded as descriptions of high-level tasks, and vice versa. As we pointed out in Chapter 1, the word "task" is often understood as actions intended to achieve certain goals.)

A second key difference between HTA and GDTA is that in HTA it is assumed that higher-level goals are typically achieved through teamwork and lower level goals through individual work (Annett, 2000, p.34). In GDTA, the focus is generally on the individual worker, "... determining what aspects of the situation are important for a particular operator's situational awareness... In such analysis, the major goals of a particular job class are identified along with the major subgoals necessary for meeting each of these goals" (Endsley and Garland, 2000, pp. 148-149). This description suggests that the operator is identified first, followed by identification of the goals that are assigned to this operator. (In a third approach, which Renee Chow and her colleagues refer to as "Hierarchical Goals Analysis," the first step is to identify and decompose goals for the entire system, before any goal is assigned to any one decision maker, team, etc.; Chow, 2007).

Unlike some forms of task analysis, GDTA does not assume that tasks can always be defined as strictly sequential or linear sequences actions. It does not assume that jobs can be defined as a lock-step series of procedures or even as hierarchies of branching dependencies. Rather, GDTA takes as its starting point the fact that in complex cognitive systems, situation awareness involves a constant juggling back and forth between multiple and sometimes conflicting goals, on the one hand, and the processing of information in ongoing situations, on the other hand. In other words, the goals people work toward, and the action sequence alternatives they chose, are dependent on context.

In a GDTA, the major goals of a particular job class are identified first. When one asks domain practitioners what one of their main responsibilities is and what their immediate goals are in conducting it, the reply is often couched in terms of the technologies with which they have to work, the "environmental constraints" on performance (Vicente, 2000). Thus, as a hypothetical instance in weather forecasting, the practitioner might say,

> Well, I have to determine the valid interval of the model initialization, but to do that I have to access the last model run using the AWPIS system here, and then compare that to the following model run's initialization. Things might have gotten tweaked or biased. . .

At that point the analysts interjects: "No, what is it that you *really* are trying to accomplish?" and it invariably turns out that the "true work" that has to be accomplished falls at a more meaningful level, the knowledge level if you will, perhaps something like,

> Well, I want to know is COAMPS is the preferred model of the day or if the ensemble models are beating it up. That will tell me how much to trust the forecast low here in the southwest.

It is this clear focus on the meanings of goals and task activities that perhaps distinguish GDTA analysis from some other forms of cognitive task analysis. An example GDTA diagram appears in Figure A.2.

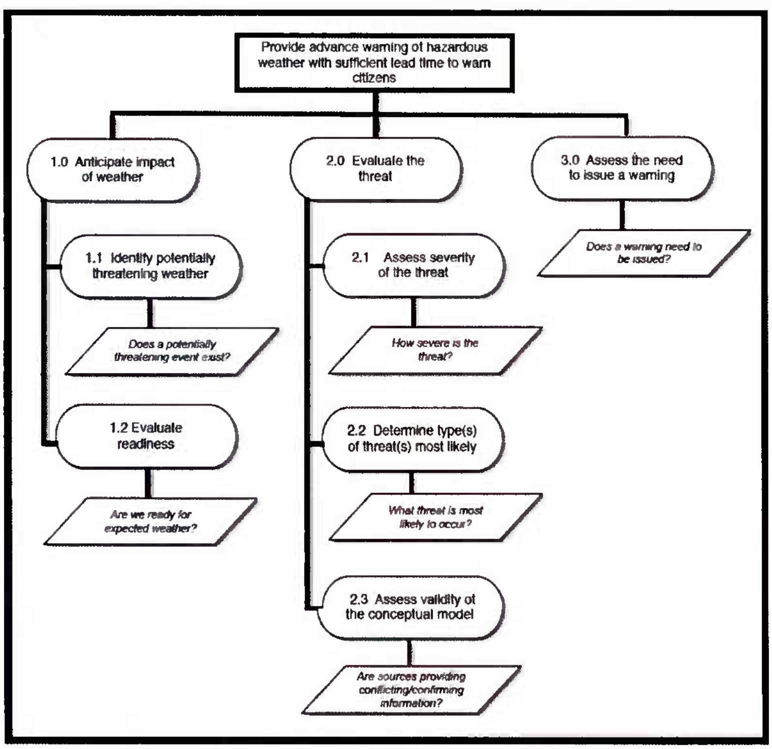


Figure A.2. An example of a GDTA diagram.
(Courtesy of Mica Endsley, SA Technologies, Inc.)

The analysis then moves on to specify the major subgoals necessary for meeting each of the major goals. The major decisions that need to be made are identified for each subgoal. Also specified are the information requirements for decision making—the information needed for the human to maintain good "situational awareness." Subgoal requirements often focus not only what data the human needs, but also on how that information is integrated or combined to address each decision. This provides a basis for determining what meanings the operator needs to derive from the data. In the Figure A.2 example, information requirements for subgoal 1.1 (identify potentially threatening weather) would include the location of high and low pressure regions, the energy available for convection, winds at various heights in the atmosphere, evidence of lifting (sea breeze, surface heating), and so on.

One sees here that GDTA is similar to Hierarchical Task Analysis, which was developed in task analysis and Human Factors from the 1960s through the 1980s. The advance over "traditional" task analysis represented by HTA was the realization by Human Factors psychologists that new jobs created by new technologies could not be described as liner sequences of activities. Rather, there are contextual dependencies and choice points—which mean that one needs hierarchical representations that describe task branchings, or stated in another way, goal/subgoal relationships (Drury, et al., 1987).

One way of understanding GDTA is to see it as a method that is complementary to the CDM procedure. In the CDM the domain practitioner is guided in retrospecting about particular past experiences, previously encountered tough cases. In the GDTA the practitioner is guided in discussing their goals in a generic sense, not necessarily tied to particular experiences or past cases. The GDTA does not seek to generate an event sequence according to a timeline. Nor does it attempt to capture invariant goal priorities, since priorities change dynamically and vary across situations. In the CDM the practitioner will almost of necessity describe previous cases in terms of the actual work that was performed using the tools and technologies that were available. But in both the CDM and GDTA seek to describe the "true work" divorced from particular technologies. For instance, one might ask a weather forecaster:

> *"What do you have to do when thunderstorms are approaching?"*

At which point the practitioner might say,

> *"I have to take data from the Warnings and Alerts System and reformat it to input it into the Pilot Information System because the data fields are different and have different parameters"*

At which point the interviewer would say,

> *"No, what are your goals, what do you have to accomplish?"*

And at that point the practitioner might say,

> *"Well, I basically have to let the pilots know that they will be entering rough weather."*

The idea is to cut through the actual work and perceive the true work in terms of the goals that have to be accomplished. The GDTA focuses on what information decisions makers would ideally like to know to meet each goal, even if that information is not available given current technology. Additionally, the particular means a practitioner uses to acquire information are not the focus since methods for acquiring information can vary from person to person, from system to system, from time to time, and with advances in technology. Once this information has been identified, current technology can be evaluated to determine how well it meets these needs, and future technologies can be envisioned that might better take these needs into account.

NDM research has contributed to theory as well as to method.

Theoretical Contributions of NDM

NDM has advanced many ideas about cognition and reasoning. We describe two main theoretical contributions. One extends the classical psychological notion of recognition into the analysis of cognitive work and the other extends the classical psychological notion of attention.

From the Notion of Recognition to An Integrated Model of Proficient Reasoning: Recognition-Primed Decision Making

As we explained above, the normative decision-analytic model of decision-making became a target for criticism because in domains involving time-pressure it is impossible for the decision maker to conduct a procedure expressing the costs and benefits of all the alternative courses of action (cf. Orasanu and Connolly, 1993; Beach and Lipshitz, 1993; Cohen, 1993a,b). Through the 1980s a number of researchers in the NDM movement developed new models of decision-making (for a review, see Lipshitz, 1993). A number of researchers who had been studying decision-making in applied contexts for years (e.g., Ken Hammond, Jens Rasmussen) had also concocted models that were embraced by the NDM paradigm (Hammond, 1993; Rasmussen, 1993). These models converge in a number of respects (Lipshitz, 1993). They were all based on:

- Appreciation for the fact that decision-making in real-world contexts is not a single process, but comes in a variety of forms involving differing strategies and differing sequences of mental operations,
- Appreciation for the effects of context and the important role of situation assessment in problem-solving in real-world situations,
- Appreciation of the role of mental simulation in the medium of mental imagery, or what cognitive scientists were calling "mental modeling,"
- Rejection of the notion that real-world decision-making culminates in a particularly critical event that can be isolated and called "the decision point," and,
- The belief that prescriptions for effective problem-solving and effective support for decision-making come not from formal analytical idealizations but rather from a solid empirical descriptive base that comes from field research, including studies of experts, rather than the traditional academic laboratory.

Dovetailing with research on decision-making under time pressure (e.g., Payne, Bettman, and Johnson, 1988; Zakay and Wooler, 1984), Klein, et al. found that most of the critical decisions were made within less than a minute from the time that important cues or information became available. (All of the longer decisions were for cases where the fire emergency itself lasted for days). But the most striking finding was:

> [H]ow rarely we found any evidence that the fireground commanders attempted to compare or evaluate alternatives at all. In only 19% of the decisions was there evidence of conscious and deliberated selection of one alternative from several. (Almost half of these were from an incident where [experience] was low and time pressure minimal)... Most commonly, the fireground commanders claimed that they simply recognized the situation as an example of something they had encountered many times before and acted without conscious awareness of making choices at all. Phrases such as "I just did it based on experience," and "It was automatic" were the most frequently encountered (Klein, 1989, p.2).

The experts were probed repeatedly about alternative options, to no avail: "Look, we don't have time for that kind of mental gymnastics out there. If you have to think about it, it's too late" (expert quoted in Klein, 1989, p.3). The experts seemed to make decisions on the basis of a process of matching a current situation to a course of action. Sometimes this could be expressed in terms of a comparison to previously encountered situations, but only incidentally on analogy to *particular* past cases (e.g., a fire involving a billboard on a rooftop brought to mind a past case involving a billboard). Rather, matching seemed to be to what might be called memory schemas or prototypical cases, and when a given situation departed from the typical, the expert's situation assessment changed and there was a change of plan. This had strong implications, for it suggested that problem solving and decision-making are not always distinct or separate activities, as implied (if not mandated) by the traditional decision-analytic model, as depicted in Figure 9.2 above. Furthermore, it suggested that real-world decision-making is not a process of "optimizing" or finding the best solution, but a process of "satisficing" or rapidly finding an effective solution (after Simon, 1955).

Klein, et al. referred to this as "recognition-primed decision-making" (RPD). According to the RPD model, the decision maker spends most of his/her time evaluating situations rather than evaluating options. Acceptable courses of action are determined without conscious deliberation and evaluation of alternatives (or at least, they are comprehended very rapidly). Commitments are made to courses of action even though alternative courses of action may exist. Experts performing under time pressure rarely report considering more than one option. Instead, their ability to maintain situation awareness provides the decision-maker the important cues, provides an understanding of the causal dynamics associated with a decision problem, and directly suggests a promising course of action, which in turn generates expectancies (Klein, Calderwood, and MacGregor, 1989, p.463).

The initial RPD model is depicted in Figure A.3.

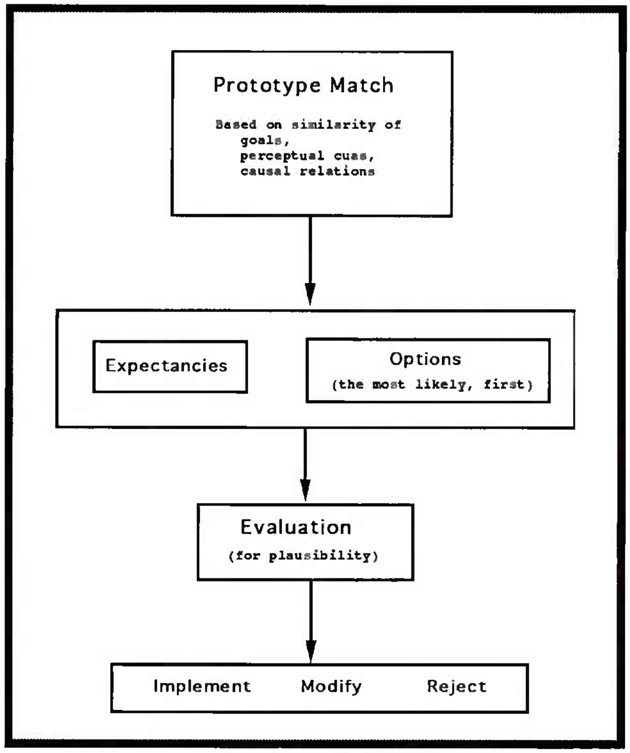


Figure A.3. The initial RPD model.

According to Klein and his colleagues, the RPD model seemed to apply not only to their own results, but results of other research on decision-making under time pressure. For example, even the classic on expertise in chess (De Groot, 1978)—regarded as a classic in cognitive science—converged with the criticisms of the decision-analytic model. Chess masters, when confronted with game boards, usually, and rapidly, identify the best move or strategy as the first one they think of; whereas novices are more likely unable to generate the best options, let alone generate them first (Klein, Wolf, Militello, and Zsambok, 1995). Studies on what happens when a decision-analytic strategy is induced and the time pressure is brought to bear showed that the strategy breaks down, and gives way to what some have called a more "intuitive" approach (Howell, 1984; Zakay and Wooler, 1984; see also Hammond, Hamm, Grassia, and Pearson, 1987). In a study of urban firefighters (Klein, et al., 1987), both more (11 years) and less (1 year) experienced fireground commanders participated. Results from the CDM showed that:

- Both groups relied heavily on situation assessment and RPD, but it was the less experienced practitioners who were more likely to deliberate over and evaluate alternative options or courses of action.
- The recognition-priming strategy was more frequently utilized by the more proficient practitioners.
- The recognition-priming strategy was more frequent even at non-routine decision points—places where one would expect it to be more likely that one might find evidence of concurrent evaluation, and,

- When experts did engage in deliberation, it was more likely to involve the deliberation of alternative situation assessments rather than alternative options.

The finding that less experienced decision makers were more likely to engage in an analytic process was striking since it had been hypothesized (e.g., Beach and Mitchell, 1978; Hammond, Hamm, Grassia, and Pearson, 1984) that people with more experience would be *more* likely to rely on a decision-analytic strategy. So it seemed that the decision-analytic model is less a model of actual expert problem solving than a model of apprentice (or junior journeyman) problem solving. If one is a novice, apprentice or junior journeyman in a task domain (that is, if one is still largely still "in-training"), then there is no way he or she can engage in a recognitional strategy. Some found these findings and conclusions to be disconcerting:

> Recognitional models may not be as satisfying as analytic models. One of the difficulties of coming to grips with a recognitional model is that so much of the important work is done out of conscious control. We are not able to become aware of how we access the memories or recognize patterns. This can be frustrating for applied researchers who would like to teach better decision-making by scrutinizing and improving each aspect of the process. In contrast, analytic models of decision-making offered the promise of bringing into the open the major tasks of evaluating options. The only thing hidden was the generation of the options themselves, and the antidote here was to generate them as exhaustively as possible (Klein, 1989, p.59).

Klein (1989, 1993) described the human, domain, and problem features that would support, induce, or even require a decision-analytic or concurrent evaluation strategy, as opposed to an RPD strategy:

- Less experienced decision makers,
- Unfamiliar tasks,
- Tasks involving data that are abstract or alphanumeric,
- Tasks that include steps that depend upon computation or involve a mandated formal analysis or step-wise procedure,
- The presence of conflict over how the situation, the options, or the goals are viewed,
- Little time pressure,
- The explicit requirement of optimizing the outcome,
- The explicit requirement of having to justify the decision,
- The need to reconcile conflict among individuals or groups who serve in different roles or capacities.

The research that converged on the RPD model made it clear that an outstanding problem in the analysis of expertise was (and still is) to capture and specify the process of perceptual learning, which makes possible the "immediate perception" of courses of action (Klein and Hoffman, 1993; Klein and Woods, 1993).

Subsequent research led to the refinement of the RPD model (Klein, 1997; Lipshitz and Ben Shaul, 1997). In fact, the model came to be integrated with the basic ideas about problem solving from Karl Duncker (1945) and also some new concepts that were emerging in the discipline of human factors psychology. As can be seen in the Figure 9.3 above, the initial RPD model implicitly embraced the "refinement cycle" notion from the classic model of Duncker. That is, at the terminal decision point in the RPD, the options include "modify" and "reject" in addition to "implement." If a course of action is rejected, some other course of action must be determined. Thus, one must go back and reassess the situation (in RPD terminology) or reformulate one's mental model (in the Duncker approach).

On the other hand, the initial RPD model differed from the Duncker model in that there is no place for a process in which the decision-maker takes (or can afford to take) the additional time to attempt to confirm or refute a mental model, judgment, or hypothesis. In the sorts of situations that were embraced by the initial RPD model (e.g., fire fighting), the unfolding situation yields further cues or information about the effects of an implemented action. Klein's initial research (Klein, 1982, 1987; Klein and Weitzenfield, 1982; Weitzenfeld and Klein, 1979) had focused on "comparability analysis" task in the domain of avionics engineering. In that domain it was natural for the experts to reason by analogy and comparison, that is, to solve new cases by comparison to a memory for past cases. Klein came to realize that case-based reasoning was only one possible problem-solving strategy, another important one being reasoning on the basis of a knowledge of causal relations and abstract principles—"mental modeling" (see Lipshitz and Ben Shaul, 1997). Within the span of a few years after the postulation of the initial RPD model and the refinement of the CDM procedure, Klein et al. had conducted a number of additional projects, involving over 150 CDM procedures with experts in diverse domains. The RPD model was elaborated on the basis of the findings.

One focus of the research that motivated the elaboration of the RPD was on the sampling of a variety of domains to assess the frequency with which experts relied on a recognition-priming decision strategy versus the concurrent evaluation of options. One study, for example, utilized the CDM in the study of engineers who designed simulators (Klein and Brezovic, 1986). In that study, the researchers probed 72 design decisions involving cases in which ergonomic data were needed in order to decide about tradeoffs in simulator design. Although the designers felt they were under time pressure, the decisions were actually made over a period ranging from weeks to months. Sixty percent of the decisions seemed to involve the recognition-priming strategy but 40 percent involved concurrent evaluation of options. The elaboration of the RPD appeared in different forms circa 1989, appearing in Klein (1989) and in Klein, Calderwood, and MacGregor (1989) (for a review, see Klein, 1993). The intent of the refinements was to capture differing strategies for decision-making:

1. Matching of situations to actions→
2. The development of an "action queue" when simple matching fails or when the problem situation is highly dynamic→
3. A more complex decision process in which situation assessments and action queues must be evaluated and refined.

The initial RPD model emphasized the direct or serial linking of recognition with action. However, the initial RPD model did not explicitly capture the processes involved in situation monitoring: the refinement of goals across the decision-making process, a mental operation which De Groot (1965) called "progressive deepening":

> For some cases we studied the situational recognition was straightforward, whereas for other cases it was problematic and required verification, and yet for other cases there were competing hypotheses... and these were the subject of conscious deliberation (Klein, 1989, p.8).

Another focus of research was elaboration of the RPD model to include a notion of mental model formation and refinement. In a study of the activities of nuclear power plant operators in simulated emergencies, Roth (1997) saw evidence for recognition-priming but also evidence for mental modeling, that is, the attempt of the operators to develop mental models of what was happening inside the power plan during the emergencies, supporting causal understanding, Roth referred to this as the "diagnostic and story-building" elements of decision making, The elaboration of the RPD (after Klein, 1989, 1993; Klein, Calderwood and MacGregor, 1989) incorporated a notion of "progressive deepening" as well as a path for decision analysis-like activities. This is presented in the Figure A.4.

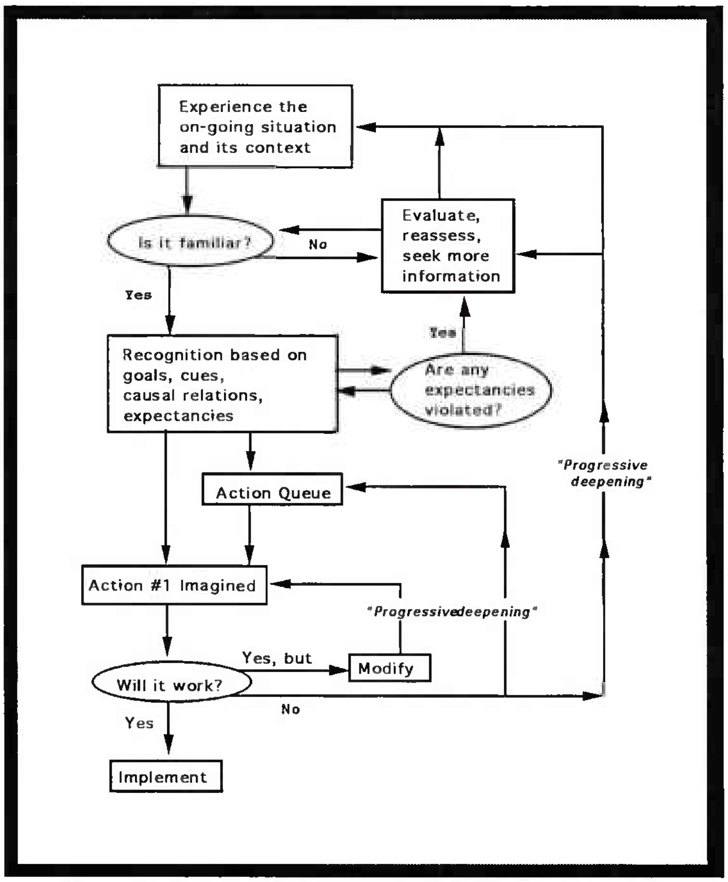


Figure A.4. An elaboration of the initial RPD model.

The "recognition" box in Figure A.4 specifies the recognition of case typicality or prototypicality in terms of cues, expectancies, and goals. Following the leftmost path straight down the model, one finds the RPD strategy. Situation assessment, an emerging concept in human factors psychology, was adopted into the refined RPD model. By hypothesis, situation assessment permits the prioritizing of cues and thereby supports selective attention, explaining why experts do not feel overwhelmed whereas novices sometimes do. Cues come primarily in the form of information revealed by on-going events. Thus, for example, in some urban fire fighting situations the commander must inspect flame or smoke color in order to make a decision or determine the timing for an action that follows from a decision.

One purpose of expectancies is to suggest ways of testing whether a situation is correctly understood through the specification of events that should occur. If expectations are violated, this can create a shift in situation assessment and a consequent refinement or shifting of goals and actions. Here, "goal" does not refer to the context-free types of goals expressed in decision-analytic models, but the specific goals and outcomes that the domain expert wants to achieve. In time-critical decision-making, goals are linked to an expectation of the timing of events. Also, the act of recognition of familiarity entails an action queue. In some situations, the action queue is a single action. In other situations the queue involves a prioritization of a set of goals or subgoals, in some situations the queue may involve the "timelining" of goals—actions may be "put on hold" pending the acquisition of additional information or the occurrence of certain events.

The action queue is potentially dynamic, as indicated by the progressive deepening loops in the elaborated RPD model. Like the initial RPD model, the elaboration of the RPD model maintained the emphasis on the fact that most decision- making occurs without conscious deliberation of alternatives. The elaboration of the RPD model preserved an emphasis on the recognition of situations in terms of familiarity or prototypicality (Klein, 1989) but takes this further by describing a process of mental simulation or "imagined action." This is clearly reminiscent of mental modeling in the Duncker model. According to Klein, Calderwood, and MacGregor (1989), the recognitional act suggests feasible goals, sensitizes the decision maker to important cues, suggests promising courses of action, and generates expectancies. But what makes the revised RPD model Dunckerian is the notion that the recognitional act makes all these things possible because it provides an understanding of the causal dynamics associated with a decision problem.

In the study of fire fighters, Klein et al. found abundant evidence for a process of imaginal simulation. In one case of emergency services, a woman had jumped or had fallen from a highway overpass, but had landed on a support strut for a highway sign. The woman was semi-conscious and the first task was to raise a ladder so that she could be held in place. Now, how should she be raised to safety? The commander considered a number of alternatives. He first imagined attaching her to a particular type of harness, but it would have to be snapped on from the back and lifting her by it would put strain on her back. He imagined using a particular type of tied strap, but that would have a similar problem. Next he imagined using a ladder belt. He imagined lifting the woman up a few inches, sliding the belt under her, tying it to her waist, buckling and snapping it, and then lifting her away from the strut. This is the option he selected, and the rescue was successfully performed (Klein, 1989). Across all the incidents that were analyzed using the CDM, "...the experienced fire fighters showed about three times more references to imagined future states, compared to the novices... [T]here is a steady increasing proportion of deliberations about situations as we move along the dimension from less experience to more experience" (Klein, 1989, pp. 36-37). One purpose of the first refinement of the RPD model was to acknowledge the role of imagery, or the mental simulation of predicted or hypothetical future states (Klein, 1989).

The emergency rescue example given above also illustrates the fact that decision-making *can* involve the consideration of more than one alternative (as in the decision-analytic model), but that the options are not generated, compared, or evaluated concurrently. Rather, they are considered and imaginally evaluated one at a time. The "evaluation" box is manifest in the first elaboration of the RPD, but the subsequent box that includes "implement," "modify," and "reject" options is replaced in the elaboration of the RPD by boxes and decisions indicating the modification and progressive deepening of plans.

Overall, the elaboration of the RPD seems to be a combination of aspects of the Duncker model (mental modeling, progressive deepening and conscious deliberation), the decision-analytic model (evaluation of alternatives under certain circumstances), and the RPD model (direct recognition leading to action plans) (Klein, 1993a; Klein and Zsambok, 1995).

The Base Model of Expertise is intended to capture a number of decision-making strategies, that is, a number of alternative sub-models can be pulled out (Klein, 1993c). In the simplest case, experience of the situation (awareness of the problem of the day and data examination) leads to recognition based on typicality or prototypicality, which leads to expectancies and an implemented course of action taken right off the top of the action queue. This sub-model, labeled as the "Recognition Priming" ellipse, reflects the initial RPD, in which a commitment is made to a course of action without any deliberation over alternatives. This simplest case may also be one of the more frequent. Looking across three of the domains that have been investigated using the CDM—urban and forest fire fighting, tank platoon commanding, critical care nursing, design engineering, etc., Klein (1989) found that expert decision-making relied on the recognition-priming more than half of the time (range of 39 to 80 percent across all of the incidents). Concurrent evaluation tended to occur only about 40 percent of the time (range of 4 to 61 percent).

In a second sub-model, situation recognition is followed by a process of evaluating the mental model and refining it prior to implementation, reflecting the Duncker model in that it involves the refinement of the initial assessment as supported by the re-inspection of data or the search for additional data. This is indicated in the "Mental Model Refinement Cycle" ellipse.

The point of the RPD is to emphasize the belief that decision-analytic procedures either cannot or simply do not occur as a part of the expert's familiar routines. However, a third sub-model of the elaborated RPD embraces the sort of situation described by the traditional decision-analytic model. In this sub-model, one goes from the initial situation assessment not to a single action that is to be implemented, but to an "action queue." Through continued looping, indicated by the "Action Plan Refinement cycle" ellipse, one can generate a set of alternative courses of action which can be assessed according to such things as costs versus benefits and likelihoods.

Neither the elaborated RPD nor the Base Model precludes the sort of analysis that is involved in the traditional decision-analytic model. Indeed, this integration regards the decision-analytic strategy, Dunckerian refinement cycles, and recognition-priming as complementary (Klein, 1989, 1993). Examples such as where to locate airports and whether to elect for cosmetic surgery might benefit from analytical strategies since the decision-maker is faced with tasks that are so new or conflict-laden that there is little opportunity for recognitional decision-making. "If a novice were to consider buying a car, the decision-analytic technique would be preferred, to systematically lay out the options and the evaluation dimensions. This would help clarify values, if nothing else. But one would not expect someone proficient, such as a used-car salesman, to go through the same exercise" (Klein, 1989, p.50).

What Klein et al. have done is to argue that each of the core concepts of a number of models that all resonate with the NDM paradigm can be embraced by a single "synthesized process model." In concert with Rasmussen's (1993) approach to cognitive systems engineering (see Chapter 10), Klein, et al. distinguish skill-based performance (i.e., recognitional skill), rule-based performance (i.e., the reliance on familiar procedures), and knowledge-based performance (i.e., reliance on conceptual principles, mental models, and conscious deliberation). The perception of typicality in the revised RPD model fits with a number of theories (e.g., Cohen, 1993) emphasizing pattern recognition. The diagnosis and situation assessment boxes are in accord with Pennington and Hastie's (1993) model emphasizing the expert's attempt to generate causal explanations.

From the Notion of Attention to an Integrated Theory of Reasoning: Situation Awareness

The concept of attention has been central to psychology from its philosophical phase through to the late 1800s when it was established as a science (Boring, 1950). Attention is at once a phenomenon of consciousness (*Of what objects or events am I aware?*), a phenomenon of perception (*What am I seeing*?), and a phenomenon of categorization (*What kind of thing or event is that?*). Attention presents to consciousness an awareness of what we are perceiving, in terms of the concepts and categories we already know (memory). It thereby allows us make judgments and decisions (Ebbinghaus, 1908, Ch. 8; Pillsbury, 1929). In traditional experimental psychology, attention was seen as a bridge between perception and memory—"What am I perceiving right now?" What one is sensing is related somehow to the concepts and categories residing in memory, allowing for perception or understanding. Numerous theories of attention have been proffered, and have been heavily researched, beginning with the earliest studies of dual-task performance or "divided attention" that helped define psychology as a science, and continuing to modern experimental psychology (Dember and Warm, 1979). Most models of attention have relied on the notion that human consciousness can only "pay" attention to a limited number of signals at a time. Attention is seen as a "filter," or as a "focusing" mechanism, or as a "limited resource" or as something that can be "captured." A number of detailed models have evolved from such metaphors, and have been refined and debated over successive generations of experimental psychology.

In a classical view of cognition (i.e., Leibniz, Descartes and many subsequent scholars), the process of sensation detects cues based on raw physical stimulus properties. These are then integrated—associations are activated and inferences are made—based upon contact with memory, resulting in meaningful percepts. Attention guides or directs this sequence. But also in classical theory, there was a subsequent process of "apperception" involving contact with the sum total of one's knowledge (cf. Ebbinghaus, 1980, Ch.12; Moore, 1939). In more modern information processing terms, this would be seen as a “bottom-up” process. Wundt (1874) referred to this as *apperceptive Verbindungen*, or "apperceptive compounds." "All recall is controlled by apperception as well as by association... Apperception selects from the possible associates those which are in accord with the entire past of the individual as well as with the single connection" (Pillsbury, 1929, p. 185). Thus for instance, a pattern of colors, shapes and movements might be detected or sensed, and with rapid contact with memory there would be the percept of "cat." Apperception would go beyond that to the sum of knowledge, and such ideas as

"I like cats" or, *"Cats are sometime seen as a symbol of evil."* This is the "assimilation of ideas by means of ideas already possessed" (De Garmo, 1895, p. 32).

But there is also a "top-down" component. Johann Friedrich Herbart, a Leibnizian associationist, introduced the notion of apperception to refer not just to assimilation but to a process whereby the contents of consciousness determine what new impressions should enter. In the language of the early foundations of educational psychology, this was a mechanism for learning and the role of learning in subsequent behavior, referred to as the "education of attention" (Pillsbury, 1926; Ribot, 1890).

The concept of attention retains its centrality, and the concept of apperception re-emerges today in applied cognitive science, in a theory of "situation awareness" (SA) developed by Mica Endsley and her colleagues (Endsley, 1995a,b, 1997; 2001; Endsley, Bolte and Jones, 2003). The point of this designation is that attention involves not just the detection of isolated signals, stimuli or cues, or even the perception of static objects, but the on-going awareness of one's environment, and especially events that one must understand (apperceive). This, in turn, supports "projection," or the anticipation of events via mental simulation. Endsley posits three levels of on-going situational awareness:

Level 1 SA concerns the meaningful interpretation of data (i.e., perception), the process that turns data into information. Hence, what constitutes information will be a function of the operator's goals and decision requirements, as well as events within the situation that is being assessed.

Level 2 SA concerns the degree to which the individual comprehends the fuller meaning of that information—a process akin to the classical notion of apperception, often referred to today as the formation of a "mental model" (see Chapter 5). In complex domains, understanding the significance of information is non-trivial. It involves integrating many pieces of interacting information, forming another higher-order of understanding, prioritized according to how it relates to achieving the goals.

Level 3 SA is the mental or imaginal projection of events into a possible future. In complex domains, the capacity to apperceive is a key to the ability to behave proactively and not just reactively. Situation awareness is critical to successful operation in dynamic domains where it is necessary for the domain practitioner (e.g., controller of an industrial process, decision maker in a military unit, etc.) to accurately perceive and then understand and project (apperceive) actions and events in the environment (Endsley, 1995a,b).

The field of Expertise Studies includes a notion of a "sensemaking mental model refinement cycle." (See Hoffman and Militello, 2008). This illustrates how there are some integrations of theoretical notions across the perspectives or communities of practice. As we show next, there are also integrations with regard to issues of design of information systems.

Implications for The Design of Information Technologies

NDM theories and research have led to approached to the design of information technologies, in once case, to support situation awareness, and in another case, to support decision making.

Situation Awareness-Oriented Design

The theory of SA has inspired an approach to the design of new information technologies referred to as Situation Awareness-Oriented Design, or SAOD (Endsley, 1995b; Endsley, Bolte and Jones, 2003). In conducting SAOD, the researcher begins with the empirical study of situation awareness. The levels of situation awareness described above form a coding scheme that is utilized in the Situation Awareness Global Assessment Technique (SAGAT; Endsley, 1987, 1988, 1990, 1995). In SAGAT, a simulation employing a system of interest (e.g., a simulation of an air traffic controller's task) is briefly halted at randomly selected times and the operator is queried as to their perceptions of the situation at that time. The system displays are blanked and the simulation is suspended while participants quickly answer questions about their current perceptions of the situation. SAGAT has been used in studies of avionics concepts, concepts for military command and control technology, and other display design and interface technologies (Endsley, 1995). The SAGAT probes are illustrated in Table A.3, which uses a scenario involving cognitive systems engineering in the aviation domain.

Table A.3. Examples of SAGAT probes, adapted for a study in air traffic control.

SA Level 1 Perception of data	What is the aircraft's call sigh? What is the aircraft's altitude?
SA Level 2 Comprehension of meaning	Which aircraft are currently conforming to their assignments? Which aircraft are experiencing weather impact?
SA Level 3 Projection into the future	Which aircraft must be handed off to another sector/facility within the next 2 minutes? Which pairs of aircraft have lost or will lose separation if they stay on their current (assigned) courses?

Research has shown that freezing can be done about a half a dozen times in a scenario trial lasting a total of 20 or so minutes, without disrupting the flow of thought. Through SAGAT, the impact of design decisions on SA can be assessed via performance, giving one a window on the quality of the integrated system design when used within the actual challenges of the operational environment. The information derived from the evaluation of design concepts can then be used to iteratively refine the system design. SAGAT provides designers with diagnostic information on not only how aware operators are of key information, but also how well they understand the significance or meaning of that information and how well they are able to 'think ahead' to project what will be happening.

Generalizing across studies, it is possible to achieve an understanding of some general principles of SAOD. For instance, the "Sacagawea Principle" (Endsley and Hoffman, 2002) asserts that human-centered computational tools need to support active organization of information, active search for information, active exploration of information, reflection on the meaning of information, and evaluation and choice among action alternatives. SAOD embodies three main steps or phases. SA requirements analysis, from SAGAT procedures and GDTA interview procedures, provides the leverage points for the design of systems to support SA. Next, SA design principles are brought to bear to translate SA requirements into ideas for system design. This process is illustrated in the Concept Map in Figure A.5.

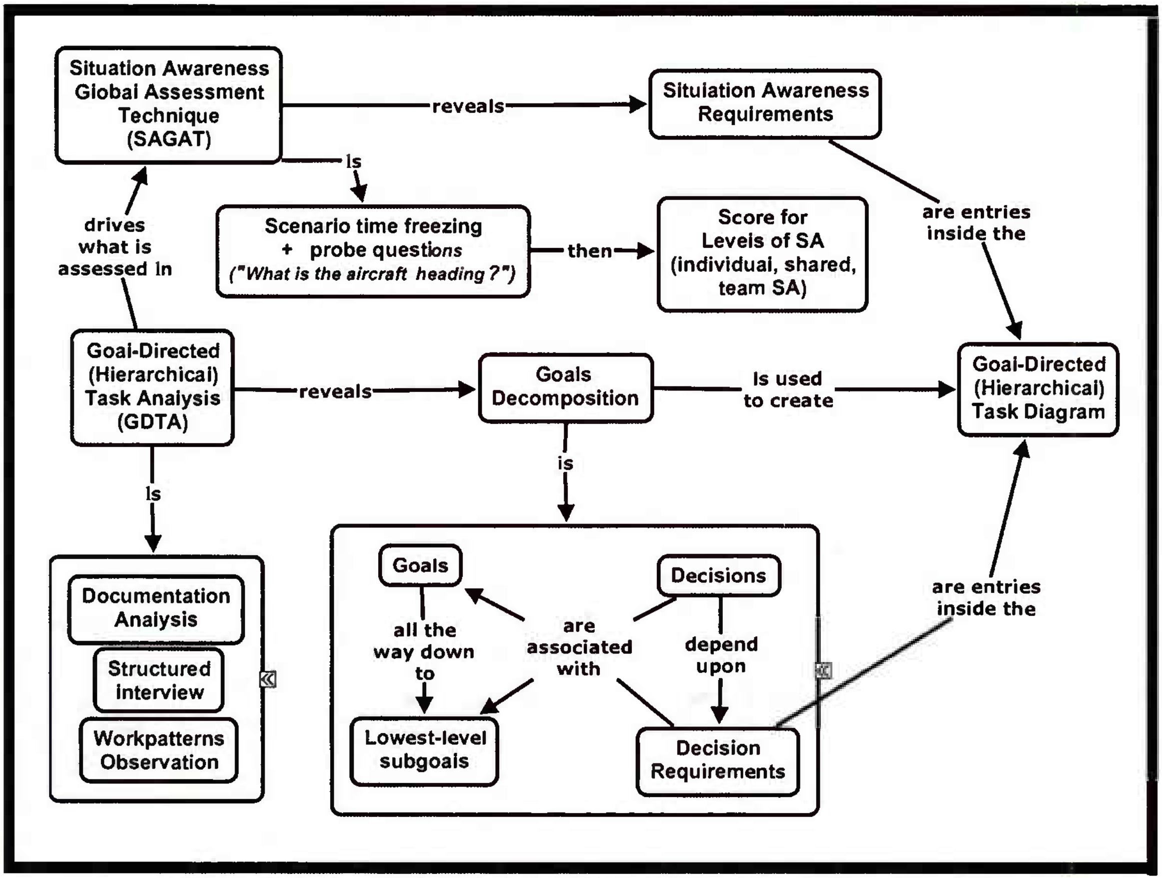


Figure A.5. How GDTA feeds into Situation Awareness-Oriented Design.

Historically, most interface design guidelines are focused at the level of the interfaces and graphical elements—fonts, how a menu should function or be placed, the best way to fill in information on a screen, etc. (Sanders and McCormick, 1992; Woodson, Tilman and Tilman, 1992). Furthermore, most guidelines assume the "one person-one machine" scenario for human-computer interaction. With regard to such cognitive processes as attention and sense-making, most guidance either remains silent on how to convey the meaning of the information that is to be displayed, or offers unhelpful generalizations (e.g., "design an interface that is intuitive" (Kommers, Grabinger, and Dunlap, 1996 p. 127). SA-oriented design principles address issues including the integration of information into knowledge and the guidance and direction of attention according to meaning, that is, how the operator manages information with the dynamically changing information needs associated with a dynamically changing situation. In some work contexts, a fixed task sequence can help constrain the layout of information. Yet for most systems flexibility is needed, allowing the operator to bounce between different goals as events change. Hence, a goal-oriented approach tries to take into account how people actually work.

A second example design principle is: *The human user of the guidance needs to be shown the guidance in a way that is organized in terms of their major goals. Information needed for each particular goal should be shown in a meaningful form, and should allow the human to directly comprehend the major decisions associated with each goal* (Endsley and Hoffman, 2002). What we might call immediately-interpretable displays need not necessarily present information in the same way it is presented in the "real world." Displays of data from measurements made by individual sensors (e.g., airspeed indicators, altimeters) may be computationally integrated into goal-relevant higher-order dynamic invariants (e.g., how a plane is flying in terms of its drag) presented as pictographic metaphors that may look fairly unlike anything in "the world" (see for example, Eskridge, Still and Hoffman, 2014). In SAOD, displays are envisioned for supporting all three levels of SA including:

1. The display of information in such a way as to support perception of its meaning with respect to goals (Level 1 SA),
2. The display of information in such a way as to support mental model formation with respect to high-level goals and possible goal conflicts (Level 2 SA), and
3. The display of information in such a way as to support mental projection and the ongoing maintenance of "global SA"—a high-level overview of the situation across all the goals (Level 3 SA).

Decision-Centered Design

Decision-Centered Design (DCD) is so named as to focus the development of technologies on supporting decision making. Its focus reflects the same motivation that is expressed in the title of the Critical Decision Method. In the evolution of the CDM, it was found that interviewing experts about their past decisions served as a good window to the identification of leverage points. Leverage points are aspects of cognitive work where an infusion of technology, however modest, might bring about a disproportionately large increase in the work effectiveness and quality. As the CDM evolved, it became clear that asking about decisions worked better than asking about other aspects of tasks, or asking experts what they know. Likewise, focusing the design of new technologies on the support of decision-making is the focus of DCD. "[Technology] should improve cognitive performance… [it should] make people smarter at what they do and their work easier to perform. Specifically, [technology] should support the cognitive activities of users and build upon and extend their domain expertise." (Stanard, Uehara, and Hutton, 2003, p. 1).

In the DCD process, the initial effort is aimed at identifying individuals who will be the "users" of the new technology, ideally the experts in the domain at hand, and coming to a rich understanding of their needs and requirements. This goes beyond typical approaches to the development of new technologies in that the understanding of user tasks and requirements is emphasized much more, and is more than just a one-off interview procedure, after which the technology developers go off and write software that is subsequently presented as a finished "deliverable."

> [Such an] initial design estimate may be based on a naïve understanding of who the user is, but as the task domain is further explored, it can become more apparent that users really need help, and that there may actually be more than one user, each with a somewhat different set of requirements to be supported (Stanard, Uehara, and Hutton, 2003, p. 2).

The DCD design approach also involves revealing and studying the really challenging and critical decision aspects of jobs. "Our working assumption is that 80% of the problems can be solved by understanding and improving the toughest 20% of the cognitive work" (Stanard, Uehara, and Hutton, 2003, p. 2). This is put in contrast with certain other approaches to the design of information technologies, such as "Situation Awareness-Oriented Design (Endsley book) and Work Analysis (e.g., Vicente, 1999), which cover broad swaths of the cognitive work that is involved in particular jobs. Thus, in DCD the probe question categories of the CDM are carried over as implications for the design of things such as interfaces: For this given problem, what are the alternative strategies, what are the critical cues and patterns that the expert perceives? What makes the problem difficult? What kinds of errors do less experienced practitioners typically make?

While NDM has generated some integrated models of cognition and cognitive work, and has spawned approaches to design, it ha recently dynamited the enterprise by invoking a distinction between models of "microcognition" and models of "macrocognition" (Klein, et al., 2003).

The Emerging Notion of "Macrocognition"

While NDM research and theory has contributed to the search for an integrated model of reasoning, the research has also dovetailed with an idea that was introduced by Erik Hollnagel and Pietro Cacciabue (Cacciabue and Hollnagel, 1995) to capture the phenomena of decision making that occur in natural settings as opposed to artificial laboratory settings.

> Although we might want to reveal specific causal sequences of various memory or attentional mechanisms, this turns out to be difficult. When we try to describe naturalistic decision making, we quickly realize that it makes little sense to concoct hypothetical information processing flow diagrams believed to represent causal sequences of mental operations, because they end up looking like spaghetti graphs (Klein, et al., 2003, p. 81).

This captures a major implication of CTA/CFR research. Specifically, it makes little sense to attempt to create any single model of how domain practitioners reason while conducting their tasks.

An example can be found in a study of weather forecasting (Hoffman, Coffey, and Ford, 2000). The procedure that was used is called the Cognitive Modeling Procedure (Hoffman, Coffey, and Carnot, 2000; Crandall, Klein, and Hoffman, 2006). The purpose of the procedure is to generate refined and behaviorally validated models of reasoning, with less effort than that involved in the method that is more often used for this application—think-aloud problem solving combined with protocol analysis.

In the first step of the procedure, each participant was presented to two alternative "bogus" models of forecaster reasoning. One or the other of the "bogus" models included the idea of mental modeling, the idea of hypothesis testing, the idea of recognition priming, and the idea of situational awareness. One of the bogus models contained a loop, the other was a linear sequence. In both models, the core concepts were expressed in domain-relevant terms. "Inspect satellite images to get the big picture" appeared in one of the bogus models instead of "Data

Examination," for instance. Furthermore, the core concepts were not arranged and linked as they are in the Base Model. Nor were they linked quite as one might expect them to be in the case of expert forecaster reasoning. For instance, in one of the bogus models, the forecaster's initial mental model was compared for agreement with one particular computer forecasting model, after which the forecast was adjusted to take local effects into account, after which the forecaster compared the forecast to another of the computer model forecasts. The intent was that the bogus models would have all the elements, using the right sort of language, would appear pertinent to the domain, and would seem not-too unreasonable but not quite right either.

The participants were asked to select the one that they felt best represented their forecasting strategy. As expected, they found this an unacceptable choice, but then they could use the various elements and ideas from the bogus models to craft their own model, one that they felt better captured their reasoning in the forecasting procedure.

Next, the researchers observed the forecasters at their jobs. Some elements of each of the reasoning models could be behaviorally validated. For example a forecaster could be observed to first inspect satellite images and other data, and then look at one or another of the computer forecasts, as they said they did when they crafted their reasoning model. Elements of the reasoning models that could not be so easily validated behaviorally were the subject of probe questions (e.g., "Did you watch the Weather Channel before you came in today?," Why are you looking at that now?"). The results showed many convergences and many differences. The notion of mental modeling was salient in all of models, especially those of the experts. It turns out that the notion of mental modeling, as we defined it in Chapter 4, is a comfortable notion to weather forecasters, especially because of the distinction, decades old, between the forecaster's conceptual understanding of a the dynamics in a given weather situation versus the outputs of the computer forecasting models.

Regarding differences and variations, models form both experts and journeymen were simple, including only some core notions, some were complex, including reference to individual computer forecast models. All of the models included what, in information processing terms, would be loops, such as the refinement cycle (e.g., if the output of a particular computer forecast model disagrees with the mental model, inspect such-and-such data and iterate until a resolution is found). In fact, all of the models included more than one loop. Some of the reasoning models had many loops or refinement cycles (as many as 7), reminiscent of "spaghetti graphs" in that everything connected to nearly everything else. Other reasoning models had just two or three loops. Some of the reasoning models showed the accommodation of local effects occurring after the formation of a mental model, some had that accommodation occurring after the inspection of satellite images ("getting the big picture") and as a part of forming a mental model. Four of The five Journeyman models, and none of the expert models, explicitly included the notion of "persistence," which is when weather dynamics are stable and a forecast can largely be a recapitulation of the previous forecast. This fits the idea that journeymen and more likely than experts to be more literal and procedure-oriented than experts.

Four months after the reasoning models had been crafted, the researchers showed all of the various models to all of the participants, with an invitation to guess the owner of each of them. The results showed that the task was confusing, in part because all of the models expressed many of the same notions, and did so in somewhat similar ways. Only 25% of the identifications were correct. As it turned out, the forecasters at this particular facility did not actually spend much time discussing their reasoning strategies with one another. Half of the participants did not correctly identify their own model. It turns out there was a reason beyond confusion. One forecaster, upon reflection, asserted that:

> "When this Bermuda High set up early a few years ago like now, the Eta and MM5 models did not handle it well but NGM did. It is the same now but we have COAMPS as well and it does well too. This model of mine does not fit my reasoning now since I am not using the [computer] models in the same way as I did when we made my [reasoning] model."

In other words, his diagram expressed a particular order of preference for examining each of the many computer model outputs in a strategy that was no longer appropriate. As Rasmussen said (1979; 1981; see Chapter 10), the practitioner does not engage in tasks, but in context-sensitive, knowledge driven choice among action sequence alternatives.

Results of this exploration in CTA methodology showed that the sequence of reasoning operations/strategies that the expert engages in is a function of the weather situation. That includes effects of oscillations that affect seasonal trends (e.g., the El Nino-La Nina oscillation, among others) and whether or not the situation is a "persistence" situation. In other words, there is nor can there be, *a* model of the reasoning of weather forecasters. To depict even all of the most typical weather situations, one would need to construct many dozens of models just for one particular region or climate. Furthermore, forecasting is always a moving target—for instance, new radar algorithms might provide a new source of data for forecasting the size of hail.

The upshot of this sort of finding for the attempt to develop models of reasoning is profound. The implication is that in real-world problem solving, mental operations are parallel and highly interacting. The description of hypothetical "basic" mental operations in such sequences as:

Attentional switching → Sensation → Memory contact → Recognition

might make sense if one is probing cognition at the millisecond level of causation (microcognition), but in the real-world context it is far more appropriate to refer to processes such as problem detection, sensemaking, re-planning, and mental simulation, which are continuous and interacting (macrocognition) and cannot be easily reduced to hypothetical building blocks placed into causal strings.

The study of micro-and macrocognition are complementary. Macrocognitive functions — detecting problems, managing uncertainty, and so forth—are typically not studied in laboratory settings. To some extent, they are emergent phenomena. No amount of research on solving puzzles such as cryptarithmetic problems or logic problems or Tower of Hanoi problems, is likely to result in inquiry about problem detection. Once these macrocognitive phenomena are identified, it is possible to trace microcognitive aspects in them. Therefore, research on microcognition is needed in parallel with macrocognition:

> We must study these types of functions and processes, even though they do not fit neatly into controlled experiments. We must find

> ways to conduct cognitive field research that can improve our understanding of the functions and processes encountered at the macrocognition level (Klein, et al., 2003, p. 83).

This suggests a new perspective on Expertise and the models presented in this report including Endsley's theory of Situation Awareness. Rather than regarding these as singular or single models that uniquely or completely capture proficient reasoning, these should be regarded as macrocognitive models, ones that attempt, perhaps with only moderate success, to capture the parallelism and interactiveness of macrocognitive functions. What they fail to capture is how variations on the models can be appropriate for particular domains, particular times, or particular local contexts, or particular proficiency levels—but that might not be the purpose of macrocognitive models. Furthermore, these are not the sorts of models that could be magically implemented in computer programs that would enable one to predict, for instance, how long in milliseconds it would take for a weather forecaster to predict fog. That is not the purpose of such models (Hoffman, Klein, and Schraagen, 2007).

Micro- and macrognition differ in a number of additional respects, and these are presented in A.4. One thing these distinctions highlight is the fundamental disconnect between the time frame for laboratory experimentation and the time frame of change in both cognitive work and the technologies used in cognitive work. Cognitive task analysis (Crandall, Klein, and Hoffman, 2006) allows one to cope with this fundamental disconnect and is ideally applicable to many of the current research challenges confronted by the DoD, broadly.

Table A.4. Distinctions between Micro- and Macrocognition.

	Microcognition	**Macrocognition**
Methodology	Controlled laboratory experimentation to isolate cause-effect relations	Field studies and cognitive task analysis Study the actual work but also reveal the nature of the true work
Methods	Traditional methods in cognitive psychology (puzzle solving, recall, recognition, reaction time) using simple, artificial tasks and materials.	Structured interviews, observations, simulations, constrained processing tasks, "tough case" tasks and other methods, using rich and realistic cases.
Participants	Typically, college students are the "subjects." They are, by definition, domain-naive. Cognition is examined in very brief experiments, looking at scales of minutes to weeks.	Experienced domain practitioners are the participants. Cognition is studied over scales ranging to entire careers. The full proficiency continuum is examined.
Ontology	Information processing/symbol system approaches assuming certain mental operations that are presumed to be fundamental or basic (e.g., short-term memory limitations)	Descriptive ontology of cognitive work: sensemaking, (re)planning, mental modeling, etc.

Phenomena studied	Typically, phenomena that generally only appear in the laboratory, such as phenomena in the solving of pre-formulated puzzles and problems.	Phenomena that are not like to occur in the laboratory and that laboratory studies would be highly unlikely to demonstrate, such as problem detection.
Explanatory Goal	Reductionist causal chain theories of cognition at the scale of milliseconds (e.g, memory access, attention shifts, etc.).	Understanding expert knowledge and reasoning, and understanding how cognition adapts to complexity. Scales can be minutes to years.
Modeling approach	Cognition is typically defined just in terms of memory and resource limitations and numerous biases (dozens have been proposed and studied). Computational models are based on parameters for processing limitations. The goal of decision aids is to mitigate bias.	Cognition is characterized by flexibility and adaptability. Bias is not typical in expert reasoning and is in many cases a laboratory artifact. The goal of decision aids is to contribute to the true work, and not proceed by imposing weak or inappropriate formal models.
Design Approach	Typically, designer-centered design in which tools and interfaces are designed with minimal user input and one or another formalism is imposed on the user (e.g., the user must input numbers to feed a Bayesian process).	Human-centered design in which tools and interfaces are based on rich user input and studies of usefulness and usability.
Applied goal	Computational models that predict performance, but performance is generally measured in superficial ways (hit rates, error rates) and performance is typically only for fixed, well-defined tasks.	Technologies that amplify and extend the human abilities to learn, know, perceive, reason, and collaborate.

The Macrocognitive perspective stands in contrast to normative choice theories of decision making. First, it is descriptive and as such its aim is to inform our understanding of what decision makers actually do rather than what they should do. Second, macrocognition posits a decision maker who is continuously engaged in monitoring the environment, reassessing the situation, and trying to understand what is going on, until decision or action is required—Macrocognition regards decision making as involving a number of interacting and parallel processes rather than seeing a discrete decision point as the culmination of a causal chain in an abstract analysis. Macrocognition does not shy away from complexity, and indeed is closely linked to systems theory and systems thinking. Thus, a prime goal of macrocognition research is to inform our understanding of resilient, robust decision making.

The implications of the micro-macrocognition distinction extend to the design of technologies: The more detailed and bounded a task is, the more likely it can be cast in stone in software, but the more likely it will be that the task description will be brittle and fleeting with time and context.

Resilience Engineering

NDM has inspired theories of macrocognitive work that identify robustness, adaptivity and resilience as ideal goals (Hoffman and Woods, 2011; Woods, 2000). But how are such aspects of work systems to be measured? Robust decision making involves more than consistency in making good decisions, it involves making good decisions under circumstances where events are unfolding, problems are emergent, and stakes are high. More even than this, robust decision making includes a capability to change the way one makes decisions, in light of novelty and emergence.

The need for measures that illuminate features and phenomena at the "systems level" is widely recognized. Traditionally, human performance is gauged in terms of efficiency measures referred to as "HEAT" measures: hits, errors, accuracy and time (Hoffman, 2010). Such measures speak to the de-humanized economics of work systems, and are blind to other significant aspects of work systems. Is the work method learnable? Does it help workers achieve expertise? Does it motivate or demotivate workers? Are the tools understandable and usable? Are the humans and machines engaged in a genuine interdependence relationship in which they can make their intent and goals observable? (see Hoffman, Hancock and Bradshaw, 2010; Hoffman, et al., 2010; Klein, et al., 2004).

The concept of "resilience engineering" has gained significant traction in the engineering and computer science disciplines (Hollnagel, Woods and Leveson, 2006). It is now a topic for symposia on resilience in cyber systems, control systems, and communication systems. Recent funded research programs include calls for the development of technologies that manifest adaptive and resilient capacities. As we have seen for many concepts that make it to the front burner, resilience may be watered down and become a mere flavor of the month through overuse and uncritical use. That is, unless a methodology is forthcoming to specify ways in which resilience might actually be measured. So, what is resilience and how can it be measured in a way that enables the creation of human-centered technologies and macrocognitive work systems?

Adaptation in macrocognitive work systems is described by five fundamental bounds (Hoffman and Woods, 2011):

- Bounded Ecology: A macrocognitive work system can never match its environment completely; there are always gaps in fitness—and fitness itself is a moving target.
- Bounded Cognizance: Limited resources and inevitable uncertainties lead to unavoidable gaps in knowledge. There is always "effort after meaning," though the struggle to acquire and deploy knowledge may temporarily ease.
- Bounded Perspectives: Any perspective both reveals and hides things, and macrocognitive work systems are limited in their ability to shift their perspective cost-

effectively. Apprehension gaps can widen because situations differ in how strongly they signal the need to shift perspectives to reveal what has been hidden.

- Bounded Responsibility: Macrocognitive work systems divide up roles and responsibilities for different subsets of goals; there are always gaps in authority and responsibility. This means that all macrocognitive work systems are simultaneously cooperative over shared goals and potentially competitive when goals conflict.
- Bounded Effectiveness: Macrocognitive work systems are restricted in the ways they can act and influence situations. Distributing activities that define progress toward goals can increase the range of effective action, but increasing the distribution of activities entails difficulty of keeping them coherent and synchronized.

These fundamental bounds serve as a way of organizing the lawful trade-offs that govern macrocognitive work systems. For example, there is a trade-off in the efficiency versus the thoroughness of plans (Hollnagel, 2009). Plans must always be made more effective and efficient, but they become cumbersome as they need to incorporate more contingencies and variations. Thoroughness expands the assessments, decisions and ambiguities, and this constrains the ability to put plans into action. This trade-off is a consequence of bounded cognizance. There is also a trade-off between optimality and resilience: Increasing the scope of the routine increases the opportunities for surprise at the boundaries. This is a consequence of bounded ecology.

Trade-off spaces can describe how cognitive systems change in response to increasing demands (tempo, cascade of effects, and the potential for bottlenecks) (Woods and Hollnagel, 2006, Ch. 9; Woods and Patterson, 2000). Trade-off functions may provide metrics for the adaptive capacity of organizations in terms of a set of parameters that characterize how organizations adapt as demands change (Woods and Wreathall, 2007).

Appendix B

List of Accepted Submissions

Andrew Abaham Dave Collins	Leeds Beckett University	Professional Judgement and Decision Making in Sport Coaching: To Jump Or Not To Jump
Simon Attfield Bob Fields Ashley wheat Rob Hutton Jim Nixon Andrew Leggatt Hannah Blackford	Middlesex University, UK TriMetis Ltd.	Distributed Sensemaking: A Case Study of Military Analysis
Chris Baber Xiuli Chen Andrew Howes	University of Birmingham	(Very) Rapid Decision Making: Framing or Filtering?
Chris Baber Gareth Conway Simon Attfield Chris Rooney Neesha Kodagoda Rick Walker	University of Birmingham DSTL Middlesex University	How Military Intelligence Personnel Collaborate on a Sense-Making Exercise
MAJ Bridgette Bell COL James Ness	U.S. Military Academy	Naturalistic Decision Making, Risk Management, and Leadership
Patrick Belling Jason Sada Paul Ward	Axon Sports, UK Huddersfieeld University, UK	Assessing Hitting Skill in Baseball using Simulated and Representative Tasks
Catherine Bolczak Amanda M. Staley	MITRE	Reducing the Information Scavenger Hunt to Improve Air Traffic Management Decision-Making
Joseph Borders	ShadowBox LLC	
Matthieu Branlat Julio C. Mateo Michael J. McCloskey Lisare Brooks Babin	361 Interactive, LLC	Revealing and Assessing Cognitive Processes Underlying Cultural Acuity Through Domain-Inspired Exercises
Wayne Chi Wei Giang	University of Toronto	
Marvin Cohen	Perceptronics Solutions,Inc.	Invited Keyonte
Françoise Darses	French Armed Forces Biomedical Research Institute	Attendee

Mustafa Demir	Arizona State University	
Beth DePass	Raytheon BBN Technologies	Attendee
Stuart Donn	British Columbia Emergency Health Services	Attendee
Peter Fadde	Southern Illinois University	From Lab to Cage: Turning the Occlusion Research Method into a Sports Training Program
Rhona Flin	University of Aberdeen	
Suzanne Geigle	MITRE	Evidence-Based Decision Making in Civilian Agencies: An Analysis of Three Cases
Matylda Gerber	Warsaw School of Economics	Intuitive Potential and Predicting Entrepreneurship – a Study on a New Method of Measuring Intuition
Julie Gore Adrian Banks Almuth McDowall	University of Surrey, UK	Advancing ACTA: Developing Socio-Cognitive Competence/Insight
Lee Green	University of Alberta	Attendee
Hana Harencarova	Institute of Experimental Psychology, Slovak Academy of Sciences	
Simon Henderson Robert Hoffman Larry Bunch Jeff Bradshaw	Cranfield University, UK IHMC, US	Applying the Principles of Magic and the Concepts of Macrocognition to Counter-Deception in Cyber Operations
Robert Hoffman Bradley Best Gary Klein	IHMC Adaptive cognitive Solutions Shadowbox LLC	
David Johnston Ben Morrison	Australian College of Applied Psychology	Exploring Cue Use in Rugby League Playmakers to Inform Training Initiatives
Helen Altman Klein Joseph Borders	Shadowbox, LLC	Good Stranger Diagnostic Tool: Measuring Capacities and Limitations to Inform Training
Gary Klein Joseph Borders Corinne Wright Emily Newsome	Shadowbox LLC	An Empirical Evaluation of the ShadowBox Training Method
Danny Shu Ming KOH Hock Guan TEE Boon Kee SOH Angela Li Sin TAN	DSO National Laboratories, Singapore	Tools for Facilitating Critical Decision Method during Tacit Knowledge Elicitation

Daphne LaDue	University of Oklahoma Center for Analysis and Prediction of Storms	Invited Speaker Climate Change
Frank Linton Sarah Bebe Mark Brown Casey Falk Mark Zimmermann	The MITRE Corporation	The SAnTA Recommender System and Naturalistic Decision Making
Jeffrey M. Bradshaw	IHMC	Invited Speaker Protecting the Nation's Food Supply
Anne-Claire Macquet Héloïse Lacouchie	French Institute of Sports (INSEP)	What is the Story Behind the Story? Two Case Studies of Decision-making under Stress
Brian M. Moon Holly C. Baxter Gary A. Klein	Perigean Technologies, LLC Shadowbox LLC	Expertise Management: Challenges for adopting Naturalistic Decision Making as a knowledge management paradigm
Ben Morrison Natalie Morrison	Australian College of Applied Psychology	Aiding Police in the Detection of Imminent Terrorist Attacks: Testing Different Approaches to Improving Situation Assessment Skills
Ben Morrison Julia Morton, Natalie Morrison	Australian College of Applied Psychology	Identifying Critical Cues in Mental Health Assessment using Naturalistic Decision-Making Techniques
Kathleen Mosier Ute Fischer	San Francisco State University Georgia Institute of Technology	The Challenges of Asynchronous Communication for Distributed Teamwork: Task Performance and Media Effects
Timothy Neville Paul M. Salmon	University of the Sunshine Coast	Look Who's Talking - In-game Communications Analysis as an Indicator of Recognition Primed Decision Making in Elite Australian Rules Football Umpires
Mark Pfaff Jill L. Drury Gary L. Klein	The MITRE Corporation	Crowdsourcing Mental Models using DESIM (Descriptive to Executable Simulation Modeling)
Georges Potworowski	University at Albany	Attendee
Olga Preveden	University of Vienna	Expertise and decision

		making in real estate appraisal: Results from a naturalistic study
Agoston Restas	National University of Public Service, Hungary	
Scott L. Rosen Jim Ramsey Christine Harvey Samar K Guharay		New Practices for Simulation Metamodeling in a Modern Computing experimentation Environment
Emilie Roth Emilie M. Roth Laura G. Militello Rhona Flin Ranjeev Mittu, Austin F. Mount-Campbell	Roth Cognitive Engineering Applied Decision Science, LLC University of Aberdeen The Ohio State University	Improving Safety: What are Leverage Points from a Macrocognition Perspective?
Ronald Scott Emilie Roth Beth Depass Jeffrey Wampler	Raytheon BBN Technologies Roth Cognitive Engineering	Externalizing Planning Constraints for More Effective Joint Human - Automation Planning
Lygia Stewart Cristan E. Anderson	UCSF and SF VA Medical Center	The Influence of Operating Room Handoffs on Teamwork, Stress, and Work: A 360 degree Evaluation of Team Shared Situation Awareness
Mr. Steve Stone	Robert Morris University, Australia	Agility in Allocating Decision-Making Rights for Cyberspace Operations
Mr. Zheng Tao	The MITRE Corporation	Applications of a Prioritization Methodology for Complex Operational Problems and Solutions
Dr. David Woods	Ohio State University	Invited Keynote Speaker

Appendix C

Meeting Papers and Posters

Professional Judgement and Decision Making in Sport Coaching: To Jump Or Not To Jump

Andy ABAHAM[a], Dave COLLINS[b],
[a] *School of Sport, Leeds Beckett University, Leeds, LS6 3QT, UK. A.K.Abraham@LeedsBeckett ac.uk*
[b] *The Institute for Coaching and Performance, University of Central Lancashire, Preston, UK*

ABSTRACT

Recently there has been increased interest in RPD to examine decision making of sport support staff, i.e. coaches. However, within as coaching where greater time is available more considered System 2 DM would also be expected. Also, given the scientific underpinnings available to coaches we would expect greater use of formalistic rules rather than substantive heuristics in the diagnostic and/or evaluative application of RPD. Against this premise 12 long jump coaches were asked to identify the strength and weaknesses of a long jump athlete and offer a view on how they would work with the athlete. All coaches were then asked to identify what they would do if their first approach didn't work. Findings suggest that coaches have an initial wish to engage in RPD type behaviour but drawing mainly on substantive heuristics. Uncertainty pushed coaches to become more considered, and formalistic. In conclusion, coaches have the capacity to be 'expert' in their DM behaviour but may not use this capacity unless pushed to.

KEYWORDS

Formalistic rules, Substantive folk heuristics, professionalism, analytic decision making, recognition primed decision making (RPD)

INTRODUCTION

In their position paper, Kahneman and Klein (2009) agreed that decision making had the capacity to become biased and flawed through overconfident reliance on and application of heuristics to solve problems and make judgements. Such overconfidence would be borne out of thinking that a swift naturalistic judgement and decision can be made based on 'experience' when in fact a more thoughtful approach should in fact be taken. It is in this space of flawed judgement and decision making that more can be learned about coaching practice and, by association, the development of coaching practice.

Numerous researchers within coaching have identified problems of coaches making judgements drawing on 'folk pedagogy' (Abraham & Collins, 1998; Gould & Carson, 2004). The suggestion being that, while this folk pedagogy may have value, its experiential source often means it is without theoretical or critical basis. Such a position has consequences for identifying coaching practice through the lens of PJDM. If coaching is to be viewed through a professional lens then this will bring certain benchmarks with it. For example Carr (1999) has identified that professions are defined by their recourse to theoretical and/or empirical knowledge in making judgements. Furthermore, that this practice is checked, monitored and informed by a critically informed peer group. The question that arises is; does the reality match the hypothesised ideal approach? Do coaches engage in PJDM in all of their decisions? In order to understand this question it is useful to explore the system typology of put forward by Kahneman, (2011) and the recognition primed decsion making (RPD) theory suggested by Klein (2008).

Kahneman offers further useful insight, particularly about which system is used and when. For example, the vast majority of decisions are made through the Type 1 process since this is typically the most efficient in terms of using mental and time resources to solve problems and achieve goals (Kahneman & Klein, 2009). Furthermore, the Type 2 system is used less frequently since it is too inefficient (at least in the short term), slow and effortful in dealing with most day to day and moment to moment problems. In fact Kahneman states that for many people Type 2 system as 'lazy' such that "If System 1 is involved, the conclusion comes first and the arguments follow" (p. 45). This view has important consequences for defining judgement and decision making as being 'professional' as defined earlier. If coaches consistently rely on Type 1 approaches in their coaching and neglect Type 2 their capacity to be professional both as a practitioner and learner inevitably becomes compromised.

In contrast to Kahneman, Klein and colleagues own work has focused on examining how practitioners can and do make 'professional' (or expert) fast Type 1 naturalistic decisions (NDM) in pressurised circumstances; for example, fire fighting (Klein, 2008).Klein argues that professionals are able to consistently able to make correct decisions without the need to revert to slow CDM. To exemplify this capacity the Recognition Primed Decision Making (RPD) model, one of the most consistently referred to models within the NDM literature, was developed (Klein, 2008). This empirically supported model predicts that, in naturalistic environments, expert professionals are able to make use of recognized perceptual cues/patterns to make fast decisions. There are three levels to the RPD model that are enacted according to how just how recognizable the perceptual cues are. In his work examining volleyball player decision making Macquet (2009) summarised the three levels to:

1. *Simple Match.* At this level cues in the environment immediately and automatically match, with no or extremely limited conscious activity, with a decision and action.
2. *Diagnose the Situation.* This level is enacted when perceptual cues do not immediately offer a view on the expectancies in the environment. As such, the expert uses their experiential knowledge, both tacit and explicit, to simulate what may have led to the situation. A view is quickly established and that matches a course of action and a decision is made.
3. *Evaluate a Course of Action.* This level is enacted when the situation is recognized but a solution does not immediately present itself. The expert again drawing on experiential knowledge will then mentally simulate the consequences of one or two actions before choosing a course of action.

All three levels of RPD are fast acting, while only the first level is truly intuitive, as Klein states:

> The pattern matching is the intuitive part, and the mental simulation is the conscious, deliberate, and analytical part. This blend corresponds to the System1 (fast and unconscious)/System 2 (slow and deliberate) account of cognition (Klein, 2008, p.258).

Although Klein argues that this account integrates the system 2 process, there is a further argument that even here the use of system 2 is not as deliberate as perhaps it could be. An adaptation to the RPD theory was created to consider how professionals cope with uncertainty, such as when there is no immediate intuitive response available (i.e. when the 2nd or 3rd RPD processes are required). The solution, known as RAWFS, was offered by Lipshitz and Strauss (1997). These authors argue that when a professional encounters uncertainty they draw on one or more of five coping mechanisms. Four of which; *Reduce uncertainty by collecting additional information, make Assumption, Weigh up pros and cons, Forestall*[1] would align with Klein's view that professionals engage system 2. However, these and other authors identify that the use of system 2 conscious activity in these circumstances only continues until a diagnosis or action that satisfies the immediate needs of situation, or which at least buys some time, is selected – a behaviour labelled *satisficing* (Lipshitz & Strauss, 1997). Klein argues that the satisficing process is still 'expert' or 'professional' since their data identifies that this satisfying process leads to correct courses of action more often than not. This argument, however, seems to be at odds with the empirical and theoretical view of critical, theoretical and peer engaged professionalism described earlier.

In summary, the NDM view on professional practice places great emphasis on the professional's capacity to deal with issues as they arise. It relies heavily on the professional's capacity to respond intuitively, typically framing through tacit knowledge learned through experience. When intuition cannot answer the problem there is recourse to more considered problem solving. However, this problem solving is rarely fully analytical in nature since the goal is satisficing rather than optimising – bringing into question just how 'professional' the approach is or can be.

An Integrated View on DM

Of course, the NDM approach is highly valuable to those who work in emergency or military situations where a lot of Klein's work has centred. However, as pointed out by Martindale and Collins (2013), not all occupations, are defined by such high-pressure, short time frame environments. Sport professions such as coaching and sport psychology (Abraham, Collins, & Martindale, 2006; Martindale & Collins, 2012) would still be identified as 'naturalistic' yet may well benefit from spending more analytical time (Yates & Tschirhart, 2006) on problems as opposed to simply satisficing. In fact, for all these professions critical thinking, planning and reflective practice are seen as being crucial to effective practice (Knowles & Gilbourne, 2010; Strean, Senecal, Howlett, & Burgess, 1997). Indeed the simplistic, yet not completely unrealistic, view of coaching being a Plan-Do-Review process would suggest that two major parts of the process have the potential to *not* be time pressured. For example, Schön (1991) refers to the importance of both reflection *on* as well as *in* practice (in practice

[1] The underlined capital letters spelling RAWF. The missing S relates to a 5th option, which is to simply Suppress uncertainty.

presumably being similar to the more thoughtful aspect of RPD) for informing and developing professional practice. However, even though coaches (and other sport professionals) typically do have more time available to them than a soldier in a combat setting, there will be times when quicker decisions need to be made in training (i.e. intervening in a practice) or competition (half time team talk). So how does one retain a professional status in naturalistic settings if a fully analytical DM is not possible? Is PJDM possible in naturalistic settings? The answer to this question must be in the way that the Type 1 and Type 2 processes talk to each other.

An insight to answering the question of professionalism comes from the review of DM and judgement by Yates and Tschirhart (2006). Among a broad range of issues covered by these authors they suggest viewing DM as being an opportunity to engage in:

- *Full analytical DM.* This strongly relates to the analytical Type 2 DM suggested by Kahneman (2003).
- *Rule based DM.* This strongly relates to the heuristic based DM identified by Kahneman (2003) and the Diagnose and Evaluate options within RPD identified earlier.
- *Automatic/intuitive* DM. This strongly relates to the Type 1 ideas of Kahneman, (2003) and the Simple Match option of RPD.

Notably, however, Yates and Tschirhart (2006) augment their view on decision making with a view on the judgment that precedes it. They provide a distinction of how analytic and/or rule based decision making may follow a *Formalistic* or *Substantive* to problem solving, making judgements and therefore making a decision. They identify that formalistic judgment draws on established formal 'known' rules or theory (Abraham & Collins, 1998) to guide judgement and decision making. Alternatively, they identify that substantive judgment will draw on personal theory or rules to solve problems. In other words, professional judgement and decision making should follow a formalistic path whereas 'folk' or heuristic based judgement and decision making will follow a substantive path. In short, it is theoretically possible for practitioners to maintain a professional approach, even in naturalistic settings, <u>if</u> they maintain a formalistic approach to their analytical and/or rule based judgements and DM.

Theoretical View	Summarised Description of What Happens		
Common Perception	Plan/Review	Do	
Dual Processing (Kahneman, 2003)	Type 2 Decision Making		Type 1 Decision Making
PJDM: CDM, RPD (e.g., Kahneman & Klein, 2009)	CDM		Simple Match Intuition
		Diagnose a situation and/or Evaluate a course of action	
Decision Modes (e.g., Yates & Tschirhart, 2006)	Analytic (Formalistic or Substantive)	Rule Based (Formalistic or Substantive)	Automatic/Intuitive
Reflective Practice (e.g., Schön, 1991)	Reflection On or For Action	Reflection In Action	

Table 1. A summary of the various decision making and judgement processes thought to be used in professional practice.

Reflecting these assertions, the present study aimed to explore the DM processes used by a group of experienced athletics coaches in the discipline of Long Jump when analysing, diagnosing and prescribing the needs of a single long jump athlete. Furthermore, drawing on Yates and Tschirhart's (2006) view that "people resort to formalistic procedures only when they can't use substantive ones, which are much more natural" (p.433) the study also aimed to explore what coaches would do when presented with uncertainty regarding their judgements. In taking this approach the following research questions were developed:

1. What approaches to DM do coaches take when presented with a contextualised real world coaching problem?
    a. What knowledge source do they draw on?
2. How do coaches respond when placed in position of uncertainty?
    a. What knowledge source do they draw on?
3. What conclusions can be drawn regarding the identification, measurement and evaluation of coaching practice?

METHODS

Participants

Participants were 12 British and Irish athletics coaches (all male; mean age 43.2, sd =3.6; mean years coaching 11.2, sd= 3.8), recruited by personal contact. All had coached athletes to at least national level (participation of at least one athlete in at least one national championships) in a horizontal jumps event. At the time of the investigation, all were actively coaching. All participants were assured of confidentiality and provided informed consent.

Procedures

Participants were presented with film (8 jumps at various venues and of various distances) plus competitive records and training data on a "US varsity level" long jumper, age 20 and with a Personal Best (PB) of 8.05. In fact, the stimulus was a conglomerate of several similar North American athletes, assembled in consultation with two NCAA Division I athletics coaches to generate a consistent picture of a "good, up and coming athlete", based on the standards prevailing at that time.

All participants received the information pack at least five days in advance. They were then interviewed in a single data collection session (lasting between 45 and 70 minutes) covering two stages. Under the first, participants were asked to describe:

- Their evaluations of the athlete's strengths and weaknesses
- Their main aims for his immediate future development
- Some exemplar activities which you would employ

Participants were also asked to present a rationale justifying their decisions.

In the second stage and in order to introduce the element of uncertainty, participants were told to imagine that this diagnosis and treatment was not working and to reconsider what else they would do, using the same structure as in the first scenario. At this stage, two participants observed that this "simply wouldn't happen" and refused to complete the second scenario. Both were removed from the investigation.

Data analysis and member checking

Data were transcribed and analysed using inductive analysis (Côté, Salmela, Baria, & Russell, 1993) by a highly qualified athletics coach and experienced coach educator who was familiar with the sport and the event. Drawing on this inductive analysis a *knowledge audit* (this looks to capture key aspects of expertise) was completed creating a *cognitive demands table* (a means of synthesising data) elements of Applied Cognitive Task Analysis (Gore & McAndrew, 2009). Finally, the responses and decisions from the coaches initial responses were deductively aligned against the approaches identified in table 1. Additionally, the resonses from the second stage of the interview were also deductively aligned against the RAWF model.

RESULTS AND DISCUSSION

Against the purposes of the investigation, results are presented focused on the perceptions, intended actions and reasoning reported within a cognitive demands table previously identified. Results from the ten participants who completed the whole investigation are presented in Tables 2 and 3. In all cases, the primary reasons and actions reported by a representive sample of 5 participants[2] coach are presented; that is, the one they and the analysing coach felt was the most important rather than the one which they said first. Aligned with these responses, a deductive view on the approaches to problem solving and decision making used by the coaches are presented in the final column.

Reflecting the expected application of NDM style approaches in the first instance, participant responses in table 1 display a personally orientated substantive approach. Our deductive alignment of response to substantive as opposed to formalistic is made on the basis of the intuitive application of heuristic problem solving procedures to both diagnose and evaluate their course of action. For example, justifications for the diagnosis made and the actions suggested are almost all exclusively grounded in "my experience tells me..." and "this looks like when...." style explanations. Perceptions on strength and weaknesses, and planned actions, reflected the initial snap diagnosis with an expected response being their evaluation. There was some similarity between the coaches, resulting in some level of clustering, i.e. those who thought the issues for the athlete were technical whereas others thought the issue was one of strength and conditioning. However, the results in table 2 are probably more defined by their apparent inter-individual variability depending on their initial evaluation. In short, we suggest that responses were personally and substantively orientated, based almost exclusively on the coach's immediate intuitive perceptions and application of athletic folk heuristics.

Interestingly, when pressured by the manipulations and placed in a position of uncertainty by suggesting that their initial diagnoses/plans were not working or even incorrect, participants spontaneously assumed (i.e. Assumption based reasoning from RAWFS referred to earlier) a "back to basics" approach (see table 3). This approach was almost identical across coaches and reflected a greater reference to a more formalistic knowledge that was, apparently, aligned with deterministic modelling identified as being required for an detailed view on key components of the long jump and the role of focusing on the take-off (Graham-Smith & Lees, 2005).

---

[2] Simply a space saving measure, all results can be made available

Notably, the response to the uncertainty manipulation resulted in all coaches talking about the need to reduce uncertainty by acquiring more information, as coach 2 said, "I'll need to take a longer slower look at the key parts of the event". (Coach 2, table 3)

This more thoughtful analytic approach was also supplemented by a strong desire to get the opinions of other coaches to support the diagnostic view; "Checking with other coaches also helps to check that you are on the right track" (Coach 3, table 3) "If in doubt watch some more, usefully with another coach and a camera" (Coach 6, table 3)
Of further note was that only Coach 8 stayed with his original diagnosis, although accepting that what he had done must be at fault if no improvements had taken place. This is of note since this was the only coach who seemed to engage a more formalistic needs analysis approach in his response to the first stage of the method.

Coach	Perceived athlete profile	Rationale	Aims and actions	Rationale	Deductively Aligned DM Approach
1	"Very powerful, good speed"	"He's like my athlete XXXX. Similar flat speed figures, just jumping further"	"I'd like to work on his attack at the board ..get more of that power translated into distance."	"That was what worked for XXX. He really benefitted from that focus. This guy is very similar."	NDM – Intuitive Diagnose Draws on Substantive knowledge
2	"I like this guy's consistency. He has a good rhythm on the run-up. He doesn't seem to foul much."	"In my experience, getting the run-up right is the most important factor. So long as he's powerful enough, everything else will follow."	"Get him in the gym more. He looks the part but I would like to get his power up so he can work his technique to best advantage."	"Once you've got the consistent technique, it's all about how much power you can put down."	NDM – Intuitive Diagnose Draws on Substantive knowledge
3	"Needs even more speed....pure and simple"	He reminds me of YYYY (*coach's former athlete*). A strong boy but we just need to get him faster on the runway."	"A hard winter working on speed should do it. Whenever I take on an almost mature athlete, that's always my first action."	"I've always had success with this method. I expect it to work here as well."	NDM – Intuitive Diagnose Draws on Substantive knowledge
4	"A focus on his running mechanics. He needs to be quicker and smoother on the approach."	"My experience in biomechanics tells me by eye that the approach is this athlete's weakness."	"Use of video feedback as we work on his technique."	"As I said before, it's the approach I use."	NDM – Intuitive Diagnose Draws on Substantive knowledge. Some evidence of recourse to formalistic knowledge
5	"Greater core strength. He looks like he folds a bit on take-off so all his speed isn't converted."	"Conditioning is paramount for this event. In my experience, you cannot neglect this."	"Hard work through the winter....miss the indoors and push for a stronger athlete into next summer's events."	"I've found that they take a while to convert to my ways of thinking. Going for an indoor season is just too early."	NDM – Intuitive Diagnose Draws on Substantive knowledge. Some evidence of recourse to formalistic knowledge

Table 2. Summary of the key cognitions of five of the ten participants relating to their response to the initial stimulus asking for perceived view, aims and actions with associated rationale. The final column reflects the deductive analysis to aligned judgement and DM approach.

Coach	Perceived athlete profile	Rationale	Aims and actions	Rationale	Deductively Aligned DM Approach and Method of Coping With Uncertainty
1	"If that hasn't worked then we need to look at his contact with the board. Work on basics around the take-off."	"Most of the things I've read suggest that the event comes down to that....so we have to focus on take-off."	"So I'd still be working on his attack into the board but with more of an accuracy focus.	"All the greats are really strong at this facet. If we can get it right with this guy, it's bound to have a positive impact."	NDM – Assumption Diagnose Recourse to Formalistic knowledge Dealing with Uncertainty: R & A
2	"My next step will be to check what is happening at take-off."	"All the coaches who write about the event stress this. It's where everything works from.....or doesn't".	"A detailed breakdown of action at the board....looking for consistent trends, both good and bad."	"This is like....like back to square one. I need take a longer slower look at the key parts of the event."	NDM – Assumption Diagnose Some evidence of plans for CDM reflection Recourse to Formalistic knowledge

Coach	Perceived athlete profile	Rationale	Aims and actions	Rationale	Deductively Aligned DM Approach and Method of Coping With Uncertainty
					Dealing with Uncertainty: R & A
3	"Well if making him quicker isn't transferring into performance, we need to go back to the take-off."	"If you look at all the great athletes, they can hit the board consistently. That's what all the books talk about."	"Let's watch his last few strides, over and over, and look for trends. What is his placement, what can we tweak."	"When your ideas don't work, its back to basics. Checking with other coaches also helps to check that you are one the right track."	NDM – Intuitive Diagnose Some evidence of plans for CDM reflection Recourse to Formalistic knowledge Dealing with Uncertainty: R & A
4	"I would want to recheck my data. Have I got enough in the first place? Have I got the right angles and so on."	"If the initial analysis is not working then we need to check back, in slower time."	"If we can get slow motion at the board, that would probably unlock the solution."	"A second, more careful evaluation. Make sure we got all the relevant points."	NDM – Assumption Diagnose Some evidence of plans for CDM reflection Recourse to Formalistic knowledge Dealing with Uncertainty: R, A & W
5	"If it isn't core strength then it is certainly something at the board".	"Whenever us coaches get together, we always talk about what happening at take-off. That seems to be a consistent idea."	"I would want to get some external views on this...some filming and analysis, some other opinions."	"If my approach isn't working, it is surely sensible to get some others at the problem."	Some suggestion of CDM NDM – Intuitive Diagnose Recourse to Formalistic knowledge Dealing with Uncertainty: R, A & W

Table 3 Summary of the key cognitions of five of the ten participants relating to their response to the secondary stimulus when uncertainty introduced but continuing to ask for perceived view, aims and actions with associated rationale. The final column reflects the deductive analysis to aligned judgement and DM approach. An additional deductive view is taken on which RAWF method is used in response to the introduction of uncertainty.

Against the review and summary of the main results offered answers to the specific research questions asked become available.

- *What approaches to DM do coaches take when presented with a contextualised real-world coaching problem?*
- *What knowledge source do they draw on?*

Evidence presented here is that the coaches' initial problem solving and decision making followed a naturalistic recognition primed response. There was some evidence that the choice of approach was intuitive, i.e. there was an immediate application of a heuristic to solve the issue that was directly attributed to 'in my experience'. However, this application was apparently to engage mental modelling that both diagnosed how the athlete had arrived at their current status (i.e. second level RPD: diagnose the situation) and created a view on how what the intervention should be. In short, there is an apparent confidence in the creating a course of action based on a diagnosis that drew on an intuitive application of mental models. Such an approach would be in keeping with work examining 'expert' performance where the conditions of a problem are recognisable and match with known interventions and ways of working.

From a knowledge source perspective, the coaches seemed to have relied on substantive problem solving heuristics to offer a view on what they were perceiving. As mentioned the views offered differed across the coaches and probably reflected 'pet' opinions and views that immediately came to mind. This would be reflective of the application of the availability heuristic as defined by Kahneman (2011). This would point directly to a lack of 'professionalism' (as previously defined) in judgement and DM and is reflective of the reality already noted by Yates and Tschirhart (2006) that people will select substantive knowledge ahead of formalistic knowledge when possible.

- *How do coaches respond when placed in position of uncertainty?*
- *What knowledge source do they draw on?*

The manipulation of introducing uncertainty in this study produced results that were in keeping with what might be predicted from the theoretical ideas offered in table 1 and 2. There was an initial assumption with what the problem might be by all but one of the coaches. This led to a strong consensus that there was a need to examine what was going on at the take off board. While only some coaches shared a view that "all the books and training would tell you to go back to the take-off" (Coach 7) the fact that this was a common theme would suggest a shared formalistic rule of how to go back to basics. Furthermore, there was an explicit identification that this recourse would lead to attempts to gain further information to further understand the problem that was occurring. Both assumptions and reducing uncertainty by collecting additional information are predicted strategies of RAWFS (Lipshitz & Strauss, 1997).

These approaches would still align with the RPD model. For example; there is an intuitive rule applied (stage 1), there is an attempt to diagnose the problem (stage 2) and to evaluate a course of action (stage 3). This explanation is consistent with Klein's view that the type 2 deliberative thinking is being engaged . However, an additional more analytical focus is suggested through more considered data collection methods, i.e. video use, and the view that discussions should occur with other coaches. In short, under this level of uncertainty the coaches wish to explore options available to them and willing to do so through checking ideas with others. This level of analysis would seem to have more to do with the analytical, deep reflections identified by Yates and Tschirhart (2006) and Schön (1991).

- *Are there any conclusions that can be drawn regarding the definition, identification, measurement and evaluation of coaching practice?*

Despite the limitations of this study, the results display that, in the context offered, these coaches engaged in judgement and decision making that matched all of ideas included in table 1. Against this evidence it would seem fair to say that in order to identify coaching practice we have to go beyond what can be observed to considering the process that led to what is observed (Collins, Burke, Martindale, & Cruickshank, 2014). However, in so doing there must be an acknowledgement that at least some of this process may be tacit and difficult to access. Furthermore, given the apparent centrality of judgement and DM to practice, this centrality must then flow through to measurement and evaluation of practice. As such, evaluation must seek to check if the quality of knowledge being used whether it is for full analytical DM or with RPD situations. This must also reflect the contexts within which judgements and decisions are made and therefore the manner in which they are made (Yates & Tschirhart, 2006).

REFERENCES

Abraham, A., & Collins, D. (1998). Examining and extending research in coach development. *Quest*, *50*, 59–79.

Abraham, A., Collins, D., & Martindale, R. (2006). The coaching schematic: Validation through expert coach consensus. *Journal of Sport Sciences*, *24*(6), 549–564. doi:10.1080/02640410500189173

Carr, D. (1999). Professional education and professional ethics. *Journal of Applied Philosophy*, *16*(1), 33–46.

Collins, D., Burke, V., Martindale, A., & Cruickshank, A. (2014). The Illusion of Competency Versus the Desirability of Expertise: Seeking a Common Standard for Support Professions in Sport. *Sports Medicine*. Retrieved from http://link.springer.com/article/10.1007/s40279-014-0251-1

Côté, J., Salmela, J. H., Baria, A., & Russell, S. J. (1993). Organizing and interpreting unstructured qualitative data. *The Sport Psychologist*, *7*, 127–137.

Culver, D. M., & Trudel, P. (2006). Cultivating coaches' communities of practice: developing the potential for learning through interactions. In R. Jones (Ed.), *The Sports Coach as Educator* (pp. 97–112). Abingdon: Routledge.

Gore, J., & McAndrew, C. (2009). Accessing expert cognition. *The Psychologist*, *22*(3), 218–219.

Gould, D., & Carson, S. (2004). FUN and Games ? *Youth Studies Australia*, *23*(1), 19–26.

Graham-Smith, P., & Lees, A. (2005). A three-dimensional kinematic analysis of the long jump take-off. *Journal of Sports Sciences*, *23*(9), 891–903. doi:10.1080/02640410400022169

Kahneman, D. (2003). A Perspective on Judgment and Choice: Mapping Bounded Rationality. *American Psychologist*, *58*(9), 697–720. doi:10.1037/0003-066X.58.9.697

Kahneman, D. (2011). *Thinking, Fast and Slow*. LondonB005MJFA2W-0-EBOK: Penguin.

Kahneman, D., & Klein, G. A. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. doi:10.1037/a0016755

Klein, G. (2008). Naturalistic decision making. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 456–460. doi:10.1518/001872008x288385

Knowles, Z., & Gilbourne, D. (2010). Aspiration , Inspiration and Illustration : Initiating Debate on Reflective Practice Writing. *The Sport Psychologist*, *24*, 504–520.

Lipshitz, R., & Strauss, O. (1997). Coping with uncertainty: A naturalistic decision-making analysis. *Organizational Behavior and Human Decision Processes*, *69*(2), 149–163. doi:10.1006/obhd.1997.2679

Macquet, a. C. (2009). Recognition Within the Decision-Making Process: A Case Study of Expert Volleyball Players. *Journal of Applied Sport Psychology*, *21*(1), 64–79. doi:10.1080/10413200802575759

Mallett, C. J., Trudel, P., Lyle, J., & Rynne, S. B. (2009). Formal vs. Informal Coach Education. *International Journal of Sports Science and Coaching*, *4*(3), 325–364. doi:10.1260/174795409789623883

Martindale, A., & Collins, D. (2012). A professional judgment and decision making case study: reflection-in-action research. *The Sport Psychologist*, *26*, 500–518. Retrieved from http://www.cabdirect.org/abstracts/20133070072.html

Martindale, A., & Collins, D. (2013). The Development of Professional Judgment and Decision Making Expertise in Applied Sport Psychology Intuitive or. *The Sport Psychologist*, *27*, 390–398.
Nelson, L. J., Cushion, C. J., & Potrac, P. (2006). Formal, nonformal and informal coach learning: A holistic conceptualisation. *International Journal of Sports Science & Coaching*, *1*(3), 247–259.
Partington, M., Cushion, C., & Harvey, S. (2014). An investigation of the effect of athletes' age on the coaching behaviours of professional top-level youth soccer coaches. *Journal of Sports Sciences*, *32*(5), 403–14. doi:10.1080/02640414.2013.835063
Schön, D. A. (1991). *The Reflective Practitioner: How Professionals Think in Action*. London: Ashgate.
Strean, W. B., Senecal, K. L., Howlett, S. G., & Burgess, J. M. (1997). Xs and Os and what the coach knows: Improving team strategy through critical thinking. *The Sport Psychologist*, *11*, 243–256.
Yates, J., & Tschirhart, M. (2006). Decision-Making Expertise. In K. A. Ericsson, N. Charness, R. R. Hoffman, & P. J. Foltovich (Eds.), *The Cambridge handbook of expertise and expert performance*. Cambridge, UK: Cambridge University Press. Retrieved from http://psycnet.apa.org/psycinfo/2006-10094-024

(Very) Rapid Decision Making: framing or filtering?

Chris BABER[a3], Xiuli CHEN[b] and Andrew HOWES[b]
[a]*School of Electronic, Electrical and Systems Engineering, University of Birmingham*
[b]*School of Computer Science, University of Birmingham*

ABSTRACT
This paper addresses two issues that arise from the challenge of studying the rapid decision making that characterises the domains of Naturalistic Decision Making. First, we are interested in *how* it is possible to make decisions in fractions of seconds. Second, we are interested in *how* such rapid decision making can be modelled. As a corollary to the first issue, we are also interested in exploring decision making which eschews the need to appeal to a concept of schema. Taking a cybernetic approach to decision making, we describe a model in which expertise is defined by the ability to filter salient features from the environment rather than in terms of the complexity of schema that is applied.

KEYWORDS
Recognition-Primed Decision Making; Schema; Mental Models; Cybernetics.

INTRODUCTION

When expert decision makers respond to a situation, they rapidly determine the most appropriate course of action. The speed of response is such that it seems unlikely that experts engage in the sort of reasoning process which form the basis of traditional decision making techniques, and thus the theories of Naturalistic Decision Making (NDM) have developed to explain how such decision making is possible. The core question is, how can someone make rapid decisions? In this paper, we propose that there is conceptual weakness in some of the dominant theories of NDM and that there is an alternative form of explanation which reflects the underlying intent of these theories while overcoming this weakness. The aim is not to overturn the NDM theories because these have proven themselves to be very useful in explaining behaviour, particularly in terms of the post-hoc accounts provided by expert decision makers, but to suggest that the initial stages of decision making might involve processes which have, to date, been under-represented in NDM theories. In short, the question is whether very rapid decision making is a matter of cognition (framing of a situation in terms of the schemata that experts develop and apply) or perception (filtering of the situation through rapid extraction of salient information).

In many NDM models, features in the environment correspond to features in schemata held by the expert decision maker which, in turn, correspond to action. This is a similar process to that assumed in the Norman and Shallice's (1986) Supervisory Attentional Control system, and can be seen in Cognitive Architectures such as Anderson and Lebiere's (1989) Atomic Components of Thought (ACT). The implication of such approaches is that experts use a schema-driven control of action. In high tempo, high stress environments (such as incident response or, indeed, many sports), the time available for a decision to be made can be defined by milliseconds (even accounting for the ability of experts to anticipate environmental states). Information is extracted from the environment and then compared to a store of schemata and then, on the basis of weighted matching, an action selected, feels as if it might involve too much cognitive activity. We argue that this high level of cognitive activity need not arise from the decision making itself but from the focus on schema (and the declarative knowledge entailed in this approach). Consequently, the question is whether it is possible to define decision making in terms of procedural knowledge? In other words, to focus less on the structure of schemata and more on the manner in which perception is tuned to the environment.

Given the manner in which NDM case studies are often (but not always) constructed through post-hoc interviews, it is not surprising that a schema-based approach could prove conceivable. Gathering these verbal reports allows concept maps to be built and it is not too difficult to imagine that the concept maps, rather than representing the information provided by the experts can actually represent the knowledge held by the experts.

[3] Corresponding Author: c.baber@bham.ac.uk; 0044 121 414 3965 ; *School of Electronic, Electrical and Systems Engineering, University of Birmingham, Birmingham. UK B15 2TT*

From this, it is an easy step to assume that, as the concept map *is* the expert's knowledge space, dealing with an incident involves activation and enactment of this concept map. Thus, we propose that many approaches to NDM not only assume that the expert decision maker approaches the situation in terms of declarative knowledge but also that decision making itself becomes a matter of negotiating the space of declarative knowledge. This further lends itself to phenomenological approaches in which expert accounts not only define the type of information that the expert is using but also define the type of decision making in which they engage. In other words, there is an assumption that the account provided by the expert after the event somehow becomes the contemporaneous account of doing decision making, rather than retrospective explanation of the consequences of these decisions. A consequence of this approach (and we propose a potential weakness) is the implication that decision making thus involves only declarative knowledge (either in terms of the repertoire of patterns held by the expert or in terms of the schema-driven search for information).

DECISION MAKING AND DECLARATIVE KNOWLEDGE

In their study of Authorised Firearms Officers (AFOs) in the UK, Mitchell and Flin (2007) suggest that the decisions to shoot or not shoot are " …likely to be influenced by the experience [and] also by their expectations from prior information" (p. 377). In this study, AFOs, in a Firearms Training System, were asked to respond to the appearance of targets when they had received a neutral (no threat) or threat briefing (indicating that the target was armed and dangerous). The briefing did not appear to have an effect on either response time or decision to shoot. Either this suggests that the decisions were *not* made on the basis of this prior information, or the prior information was not presented in a manner which could influence decision making. The authors did note that it was possible that the participants responded to cues in the scenario which influenced their decisions. In a simpler task, Luini and Marucci (2013) asked participants to respond (using key presses) to images on a screen. Comparing trained and untrained participants, they showed that response to an 'armed target' was significantly faster than to an 'unarmed target' (i.e., images with and without a gun in their hand), and that trained participants showed significantly higher correct response and significantly lower false alarms than untrained participants. Taking these studies together, we propose that the shoot-no shoot decision depends on the appropriate definition of features in the environment and we further claim that this *need not* involve recruitment of schema. Indeed, it might be the case that Mitchell and Flin (2007) could have (through its use of briefing to stimulate differences in performance) have implicitly assumed that the participants would be responding using a more elaborate and detailed schema in the threat condition. For example, the 'threat' briefing of Mitchell and Flin (2007) could be represented in the form of a concept map (figure 1).

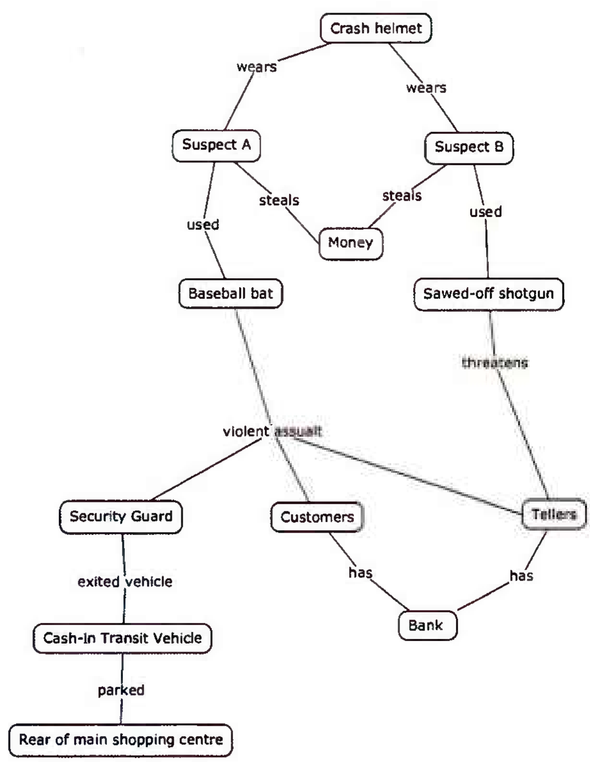


Figure 1 : Concept Map of 'Threat' scenario from Mitchell and Flimn (2007)

From the notion of a schema-driven approach to NDM, it could be hypothesised that the AFO would have some (or all) of the concept map shown in figure 1 as a 'schema', with different nodes in this schema being activated as more information becomes available. Activation of different nodes would then (somehow) activate the response options. The question is whether construction and traversal of schema can really be performed rapidly, as a schema-driven approach implies, or whether other approaches are at play.

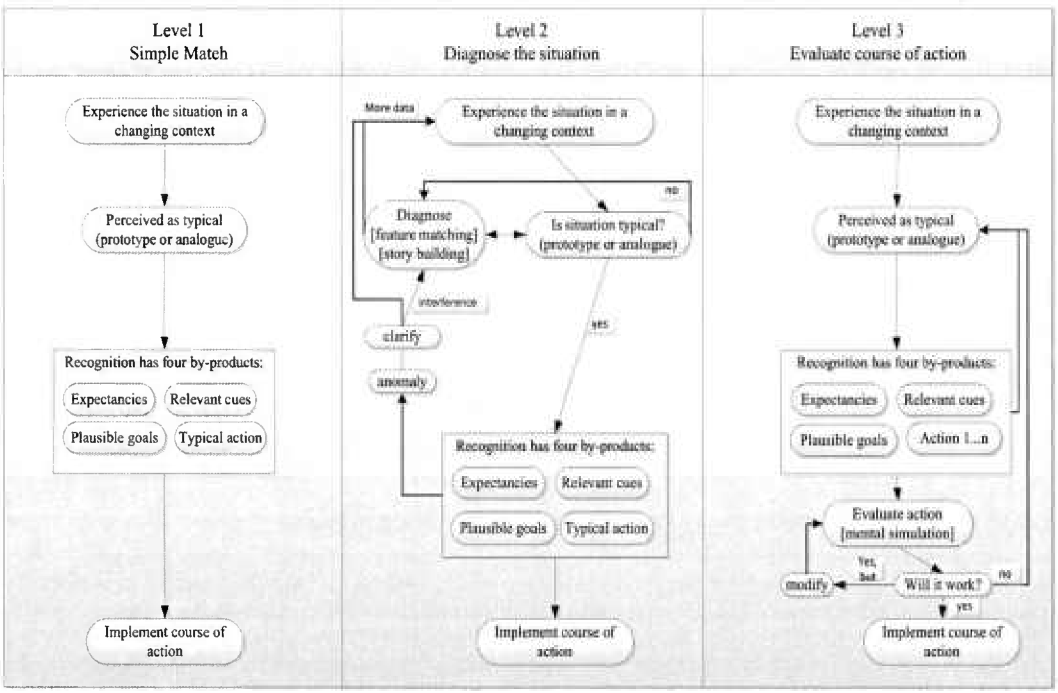


Figure 2: Recognition-Primed Decision Making

As figure 2 illustrates, RPD combines situation assessment (in terms of the decision maker experiencing a situation and determining whether or not it is 'typical') with mental simulation of responses to that situation, in order to define a plausible course of action. In terms of typicality, Klein et al. (1986) suggest that experts have a 'repertoire of patterns' based on plausible cues, goals and reactions, and this repertoire is constructed on the basis of prior experience of the expert. While figure 2 does not explicitly state *how* these repertoires might be represented, it does suggest that they involve expectancies, plausible goals, cues, and typical action. One question is whether the expert constructs this knowledge in response to the situation or whether the expert views the situation in response to this knowledge? In other words, it is possible that the expert could (on the basis of a repertoire of patterns) selectively view a situation and respond accordingly. Such an approach could comfortably fit assumptions of bias but is seldom reported in the NDM literature. This suggests that expertise is more likely to involve constructing the knowledge in response to the situation which, in turn, implies that a characteristic of expertise is not simply possession of a repertoire of patterns but also a well-practised ability to extract salient and relevant information from the situation. Thus, the concept of schema (Bartlett, 1932; Neisser, 1976), as a shorthand description of how people structure knowledge, has proved popular as a way of explaining expertise (Plant and Stanton, 2013).

Lipshitz and Ben Shaul (1997) have questioned whether we as a community are doing justice to the term 'schema' in our use of it and propose that it needs to be distinguished from the term mental model. In their study, they demonstrated that experts (in a simulated maritime combat task), in comparison with novices, collected more information before making their decisions, engaged in more efficient search, 'read' the situation more accurately, made fewer 'bad' decisions, and communicated more frequently with friendly units. They interpreted these findings in terms of Neisser's (1976) schema theory, specifically highlighting that schemata 'direct external information search', 'specify which external information will be attended to and which will be ignored', and 'become more differentiated as a function of experience' in terms of search. They go on to note that schemata 'organise information in memory' and 'direct the retrieval of information from memory', and suggest that this explains why novices (in their study) tended to repeatedly request the same information. Lipshitz and Ben Shaul (1997) distinguish schemata (which 'drive the construction of specific situation representations) from mental models (which are the product of this construction process), claiming that it is the mental model which ultimately drives the decision process because it is the mental model which contains situation-specific information structured in a way that allows coherent and consistent decisions to be made; if the mental model is incomplete, erroneous or ambiguous then decision making will be less successful. In terms of their distinction between schemata and mental models, Lipshitz and Ben Shaul (1997) seem to want their cake and eat it; if mental models are specific representations of a situation then one would expect these to involve organisation and retrieval of information from memory, which are also characteristics of schema. The only logical step (we believe) that can be taken is to claim that there is a process by which information is acquired and a separate process by which information is stored. Such a distinction is beneficial because it provides a way of separating level 1 (which is primarily perceptual) from levels 2 and 3, which become increasingly cognitive (in terms of involving more detailed and elaborate activity using mental models). The question then arises as to where action selection occurs. For Lipshitz and Ben Shaul (1997) action selection is a response to the mental model, which implies that action selection arises at level 3 (or possibly level 2). However, might it be the case that there are, to use Gibson's (1977) term, perception-action couplings, which would support action selection at level 1? Our proposal in this paper is that, rather than viewing action selection in level 1 in terms of schema, there is a simpler, more elegant description to account for the perceptual cycle that lies at the heart of expert decision making.

DECISION MAKING AND PROCEDURAL KNOWLEDGE

The approach we adopt shifts focus from declarative to procedural knowledge. To enable such a shift we adopt a cybernetic approach to human decision making, inspired by the work of Baron and Kleinman (1969). In their work, they applied concepts from control theory to model the operator of a complex dynamical system, such as an airline cockpit. In their model of cockpit instrument scanning, visual sampling is considered to occur in parallel with action selection. Sampling depends on the control task being performed.

Chen et al. (2015) demonstrate how this approach can be used to model visual search in applied and laboratory tasks. Specifically, an optimal control model embedded with the assumptions of human visual mechanisms (e.g., visual acuity degradation away from fovea, saccadic duration, and fixation duration) offers explanations for the observed human behaviours in these visual search tasks (e.g., the gaze distribution, the search time, the saccadic selectivity across colour and shape). Decision-making, skills and rules are an emergent consequence of rational adaptation to (1) the ecological structure of interaction, (2) cognitive, perceptual and motor limits (e.g., visual and/or motor constraints), and (3) the goal to maximize the reward signal. This requires a theory which allows us to predict behaviour on the basis of utility, environment and information processing mechanisms. To do this, the model uses a state representation and an optimal controller approach.

The optimal control approach predicts behaviour from a model of the temporal costs of eye and head movements, a model of how visual acuity degrades with eccentricity from the fovea, a model of cue validity, and the assumption that operators optimise speed/accuracy trade-offs. A given feature in the environment is fixated and the result of this fixation (a percept) is encoded, in terms of specific attributes. The percept updates a state vector, which is used to compare current with previous state. Thus, for example, assume that we are facing a person who might be about to use a gun. Movement of the hand could constitute a change in state. However, depending on our decision policy, movement of the hand might not be sufficient to determine whether there is a threat or not. This means that we might require further information, such as what is the person holding in that hand, before we make the decision.

Through feedback, and experience, the behaviour of the control policy comes to resemble recognition-primed decisions. The model aims to predict the operators' behaviours given theoretical assumptions about utility (e.g., a measure of the goal), psychological mechanisms (e.g., human eye-head coordination mechanism) and environment (e.g., the interaction between the operator and the interface). To achieve this goal, a state estimation and optimal control approach is used, as shown in Figure 3. In the task environment (bottom left), the operator moves head and eyes to acquire information. The state estimator (the bottom right) encodes a percept from the environment, which is then integrated with the previous state to generate a new state representation. Subsequently, the optimal controller chooses an action on the basis of the available state estimate and the current policy (which determines a state-action value function). State-action values are updated incrementally (learned) as reward and cost feedback is received from the interaction.

The control policy is a probabilistic mapping from states to action, depending on the constraints of the task and environment. This notion of a context-dependent mapping between the expert and the state of the environment feels analogous to Level 1 in figure 1. In this model, the control policy is a rule that allows the agent to select an action in terms of an action-value function. The optimal policy can be constructed by selecting the action with the highest value in each state. Using Q-learning, a form of model-free reinforcement machine learning, it is possible to define control policy as a utility function which is adjusted and tuned to the feedback from action to task performance (where task performance results in a 'reward', i.e., change in value between environment and action). It is important to note that the state representation does not rely on *a priori* assumptions about the details of specific features or their relations. In other words, the state representation makes no assumptions about the content of declarative knowledge of the person, but is focused on selecting those features which best fit the policy.

In terms of the State Representation, a state will consist of decision relevant cues. For the shooting task, the cues could involve the nature of the object in the person's hand, the posture of the person, line of sight etc. Each of these cues would have a different indication of the threat level the person presents. The state is then represented as all or some of these cues. To obtain information for these decision related cues, the model selects both eye movements and head movements (actions). The different sources of information result in different time costs and reliabilities of the information. During this process, the operators/model need to decide which cue to access, and when to stop information searching and make a decision.

In terms of Action, the output of the decision process would be to either 'shoot' or 'search for more information'. The 'search for information' would involve finding and checking an information source, which costs time to acquire and the validity of information from each source varies. This task has been studied extensively in cognitive psychology (Newell and Shanks, 2003). In terms of probabilistic inference the observations which an operator makes are noisy, and sometimes using multiple sources, and the reliability of these sources differ. In a probabilistic inference problem, the key questions concern how people integrate noisy observations, and how people weigh different sources of information. These cues can vary in the reliability of the information provided. The more cues examined, the more information gathered thus more likely to make a correct decision. However, extra time cost and/or financial cost would be required. The *probabilistic inference* task has been used in cognitive science in efforts to discover the decision-making heuristics used by people (Gigerenzer & Goldstein, 1996; Bröder & Schiffer, 2006; Rieskamp & Hoffrage, 2008).

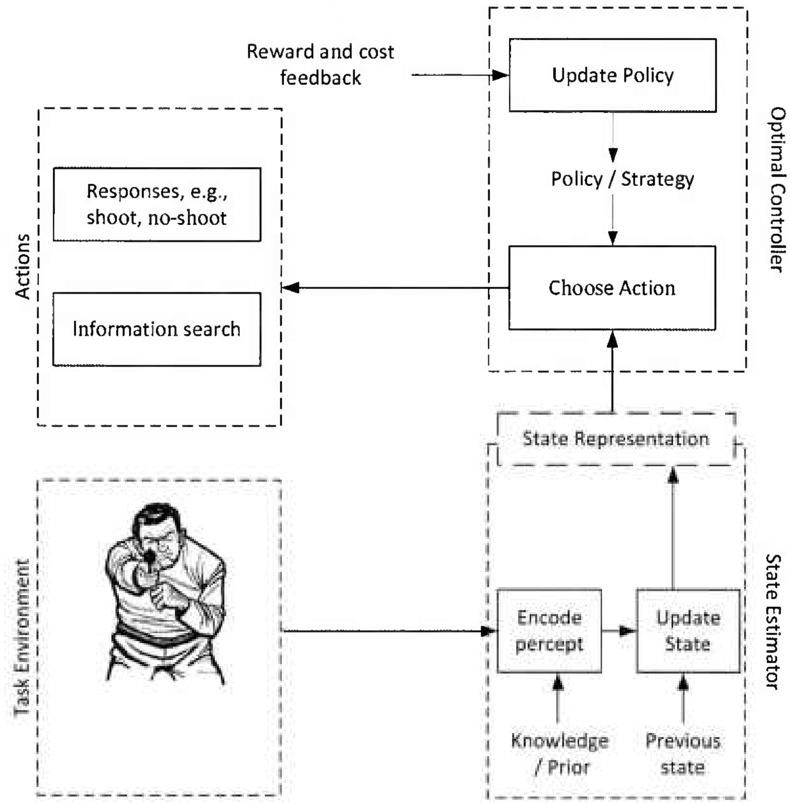


Figure 31: An overview of the optimal control model

DISCUSSION

In this paper we have considered the ways in which RPD (and related NDM models of decision making) tend to prioritise declarative knowledge (through schema and mental models) in their descriptions of rapid decision making. While such approaches might be appropriate for level 2 and level 3 decision making (in the RPD) model, which place more emphasis on the cognitive aspects of selecting options, we propose that these approaches hamper our ability to understand very rapid decision making which could be seen at level 1. To this end, we consider ways in which perception could play a key role in decision making. The argument is that the manner in which experts seek and select information becomes less a matter of managing declarative knowledge and more a matter of procedural skill and tuning. From this point of view, the expert not only has a repertoire of patterns of knowledge, experience and actions, but a set of skills which are tuned to the selection of salient information in an environment. The reasons why this distinction could be beneficial are three-fold. First, NDM is being applied to situations in which an explicit definition of declarative knowledge can be problematic, e.g., in sports. In these situations, it could make more sense to ask what features of the environment are being selected and utilised by the decision maker rather than what knowledge structure they are creating. Second, training of expertise could be supplemented by drills and practice which emphasise information selection rather than knowledge building. This is not meant to displace knowledge-based training, but to encourage thinking as to how procedural knowledge (in terms of information selection and policy weighting) could be emphasised. Third, the approach allows decision making to be modelled, which provides us with an opportunity to hypothesise strategies that decision makers might use in specific settings, and (potentially) provides an opportunity for rapprochment with 'traditional' (i.e., quantitative, optimal) approaches to decision making.

ACKNOWLEDGEMENTS

The work in this paper is partly supported by European Union project FP6-619435 SPEEDD "Scalable ProactivE Event-Driven Decision-making"

REFERENCES

Anderson, J.R. and Lebiere, C. (1998) *The Atomic Components of Thought*, Mahwah, NJ: Erlbaum

Bartlett, F.C. (1932) *Remembering : a study of experimental and social psychology*, Cambridge : Cambridge University Press.

Bröder, A., & Schiffer, S. (2006). Adaptive flexibility and maladaptive routines in selecting fast and frugal decision strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 904.

Chen, X., Bailly, G., Oulasvirta, A., Brumby, D. & Howes, A. (2015) The Emergence of Interactive Behavior: A Model of Rational Menu Search, *CHI 2015*, New York: ACM.

Gigerenzer, G. and Goldstein, D. G. (1996) Reasoning the fast and frugal way: models of bounded rationality. *Psychological review*, *103*, 650-669.

Klein, G. A., Calderwood, R. and Clinton-Cirocco, A. (1986) Rapid decision making on the fire ground, *Proceedings of the Human Factors and Ergonomics Society ??nd Annual Meeting*, Santa Monica, CA: HFES, 576-580.

Klein, G.A., (1993) Recognition-primed decision making (RPD) model of rapid decision making, In G.A. Klein, J. Orasanu, R. Calderwood and C. Zsambok (eds.) *Decision-making in Action: models and methods,* Norwood, NJ: Aplex, 138-147.

Lipshitz, R., & Ben Shaul, O. (1997). Schemata and mental models in recognition-primed decision making. *Naturalistic decision making*, 293-303.

Luini, L.P. and Marucci, F.S. (2013) Effect of emotiuons on cognitive processes : are proficient decision-makers inoculated against the influence of affective states ? *International Conference on Naturalistic Decision Making 2013,*

Mitchell, L. and Flin, R. (2007) Shooting decisions by Police Firearms Officers, *Journal of Cognitive Engineering and Decision Making, 1,* 375-390.

Neisser, U. (1976) *Cognition and Reality*, San Francisco, CA : W.H. Freeman ?

Newell, B. R. and Shanks, D. R. (2003) Take the best or look at the rest? Factors influencing "one-reason" decision making, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 53-65.

Norman, D.A. and Shallice, T. (1986) Attention to action : willed and automatic control of behaviour, In R.J. Davidson, G.E. Schartz and D. Shapiro (eds.) *Consciousness and Self-Regulation : advances in research,* New York : Plenum Press, 1-18.

Plant, K. L., & Stanton, N. A. (2013). The explanatory power of Schema Theory: theoretical foundations and future applications in Ergonomics, *Ergonomics*, *56*(1), 1-15.

Rieskamp, J., & Hoffrage, U. (2008) Inferences under time pressure: How opportunity costs affect strategy selection, *Acta psychologica*, *127*, 258-276.

How Military Intelligence Personnel Collaborate on a Sense-Making Exercise

Chris BABER[a], Adam DUNCAN[a], Simon ATTFIELD[b], Chris ROONEY[b], Neesha KODAGODA[b] and Rick WALKER[b]

[a]*School of Electronic, Electrical and Systems Engineering, University of Birmingham*
[b]*School of Science and Technology, Middlesex University*

ABSTRACT

Sense-making plays an important role in Intelligence Analysis, but can be difficult to study in situ. Thus, it is common to use training exercises to study this phenomenon. In this paper, an exercise was undertaken by Military Intelligence officers. The behaviour of groups of analysts is considered in terms the Data / Frame model of Sense-making. The paper illustrates how Intelligence Analysis need not follow a linear process but often involves parallel and overlapping explorations of data, with multiple frames that are might be minimal and sketchy. The use of representations, such as link diagrams, provides a means of externalizing frames and it is suggested that this shifts reasoning from induction to abduction as the exercise progresses.

KEYWORDS

Sense-making; Intelligence Analysis; Representations.

INTRODUCTION

While it is unlikely that there is a single, definitive way of 'doing' Intelligence Analysis, there are generic descriptions of how Intelligence Analysis could be performed. For instance, NATO (2008) describes the Intelligence (or Analysis) Cycle in terms of four phases:

- Direction: define objectives for Intelligence Requirements and Requests for Information;
- Collection: gather information by agents;
- Processing: compile and interpret information to produce intelligence product;
- Dissemination: distribute appropriate parts of the intelligence to relevant parties.

Although this implies a flow from collection to dissemination, alternative descriptions emphasize the recursive nature of the analysis process. For example, Elm et al. (2005) define this process in terms of 'down-collect' (sample from the available data for material deemed to be 'on analysis'), 'conflict and corroboration' (ensure accurate and robust interpretation of findings, and modify the 'down-collect' accordingly), and 'hypothesis exploration' (construct coherent narrative to explain the findings, and reflect this narrative back to the 'conflict and corroboration' activity). This recursion means that Intelligence Analysis is not linear (Heuer, 1999; Heuer and Pherson, 2010; Roth et al., 2010; Kang and Stasko, 2011). Such recursion is neatly captured by the Data / Frame model of sense-making model (Klein et al., 2006a, b).

Data / Frame Model of Sense-making

Central to sense-making in the Data / Frame model is the relationship between the data to which the analyst has access and the different 'frames' that can be used to interpret, make sense of, or explain, these data. Klein, Moon and Hoffman (2006a) point out that, *"When people try to make sense of events, they begin with some perspective, viewpoint, or framework – however minimal.* ***For now, let's use a metaphor and call this a frame.****"* (p. 88, emphasis added).

Most crucial of all to the Data / Frame model is the suggestion that the relationship between data and frame is both reciprocal and parallel. In other words, a frame could be applied to a set of the data, or a set of the data could suggest a frame. This reciprocity points to the continuous interweaving of activities of exploring data and generating interpretations. What is particularly useful about the notion of a frame is that it need not imply a 'solution' or final 'product' but can serve as a temporary explanatory model of aspects of the data. This accords with the suggestion from Kang and Stasko (2010) that *"...analysis is about determining how to answer a question, what to research, what to collect, and what criteria to use"* [p.25].

The point at issue is not how people use frames but how they define them in the first place (Roth et al., 2010). While the Intelligence Cycle might begin with 'Direction', this only gives a high-level sense of what the analyst might be looking for. As 'Collection' and 'Processing' progresses, new problem opportunities arise through 'discovery-led refinement' (Attfield and Blandford, 2010). Thus, one could read figure 1 in terms of a 'Direction' providing a tightly specified frame (so that the analyst will only collect and process data which are directly relevant to this frame), or in terms of a familiar problem (so the frame could be based on previous experience of similar cases), or in terms of a problem opportunity (so combinations of data would suggest particular frames which could be expanded and explained).

METHOD

The study reported in this paper use an Exercise developed for a Visual Analytics Summer School 2012 and reported at a previous NDM conference (Baber et al., 2013). Initial analysis was derived from ad hoc observation of group performance and it was felt that a more controlled approach to data collection would be beneficial. This paper presents the approach to data collection and analysis that was developed to study this Exercise.

Objective

The Exercise was designed with the assumption that the correct solution could be arrived at by defining a modus operandi (M.O.) of how the gang operated. The M.O. was as follows: a gang uses a yacht to transport drugs from Roskoff (France) to a marina in Exmouth (UK)[4]. The yacht also carries a passenger who puts the drugs into a van hired by the marina's management and drives to a warehouse in Leeds (UK). The drugs are then distributed to drivers in a mini-cab company in Leeds and sold. In order to make the exercise challenging, the data also relate to three other stories.

Procedure

The University of Birmingham ethics protocol was followed (i.e., participants were free to withdraw at any time and all data collected (including images and video) would be anonymized before reporting). Participants were given a briefing, which was intended to simulate the 'Direction' phase of the Intelligence Cycle, and the Exercise concluded with a presentation by each group, which was intended to simulate the 'Dissemination' phase of the Intelligence Cycle. The briefing was as follows:

"Muriel Grosby is a businesswoman who lives in Leeds and runs a road haulage and mini-cab firm. While she has no criminal convictions, local police have long been suspicious of her acquaintances and believe that she has links with criminal activity, particularly relating to drug smuggling and people trafficking. A known contact of Grosby, called Calabrese, was sentenced, on 14th June, to 9 years for drug smuggling.

Intelligence suggests that there is likely to be a shipment of class A drugs being delivered to a port in the South-West of England in the next few weeks. Given resource and personnel constraints, it is not possible to follow every suspect so you need to determine who should be arrested and where the best place might be to make such arrests.

Following your investigation, you will give a presentation on your findings. The presentation will include:

1. *Name of individual, or individuals, to target as Suspects.*
2. *The FIVE pieces of evidence that best support your proposal to 1.*
3. *Location of the arrest or arrests.*

In order to make this exercise easier, you will select suspects from a set of nine people:

- *Muriel Grosby, who I have already described;*
- *Jennifer Garlica who is Grosby's sister-in-law and whose husband was killed last year in what looks like a gangland hit;*
- *Vanessa Munoz who is the assistant manager of Exmouth Marina;*
- *Martina Sarti who works at the marina and is the girlfriend of the marina manager (Xavier David);*
- *Pierre Pasquidini who lives in Roskoff and travels regularly to the UK;*
- *Kenny Chiappe who drives a mini-cab in Leeds;*
- *Jake Ajachinsky who is a petty criminal;*
- *David Pico who is Jake's best friend, is also a petty criminal and has a tempestuous on-off relationship with Jake's twin sister, Denise Ajachinsky, who is also a suspect.*

For this Exercise, 'today's' date is September 10th 2012 (this will help you make sense of the dates and times on the documents you have)."

Following the briefing, participants were allocated to groups of 4-6 members[5] and then taken to their own incident rooms to complete their investigation. These rooms were equipped with whiteboards, large notepads,

---

[4] We should make very clear that the place names Exmouth, Leeds and Roskoff were included in an entirely fictional capacity and that there is no implication that any of these towns, or indeed Exmouth Marina, have been involved or implicated in any of the events in the Exercise.

[5] Contemporary approaches to Intelligence Analysis often rely on groups of people working together in 'Fusion' centers (Roberts, 2011; Treverton and Gabbard, 2008). For example, the US Army All-Source Analysis System (ASAS) involves four analysts working together to provide data for a senior analyst. We took this as a template for our study and had people working in teams of 4-6 people.

pens, post-it notes and paper. Each group was provided with a pack of 49 slides. The pack included nine suspect cards (with picture of the suspect and their correct addresses), together with a combination of telephone logs, harbourmaster logs, maps, business accounts, witness and arrest statements, newspaper articles etc. Figure 1 provides an illustration of the types of evidence supplied[6]. Each sheet of evidence contained several topics, e.g., dates, phone numbers, names, locations etc.

Figure 1 : Intelligence used in the Exercise

Participants

The study involved a workshop with serving UK Military Intelligence Officers, as part of a weekly Intelligence Analysis programme. Sixteen Staff and Officers agreed to be observed during the investigative study. The participants were divided into three groups, with two groups of five and one group of six. Five participants were female and the remainder were male.

Data Collection and Analysis

The analysis involved three forms of data: activity sampling, process analysis and review of groups' answers to the challenge. First, each group had a dedicated observer who recorded the activity on the group on an activity sampling sheet every 10 minutes. Second, each observer, when they were not completing the sampling sheet, took photographs of the diagrams that the groups were making, or of group activity, and made contemporaneous notes of the group discussions. Third, at the end of the exercise, each group presented its findings and these were recorded and analyzed.

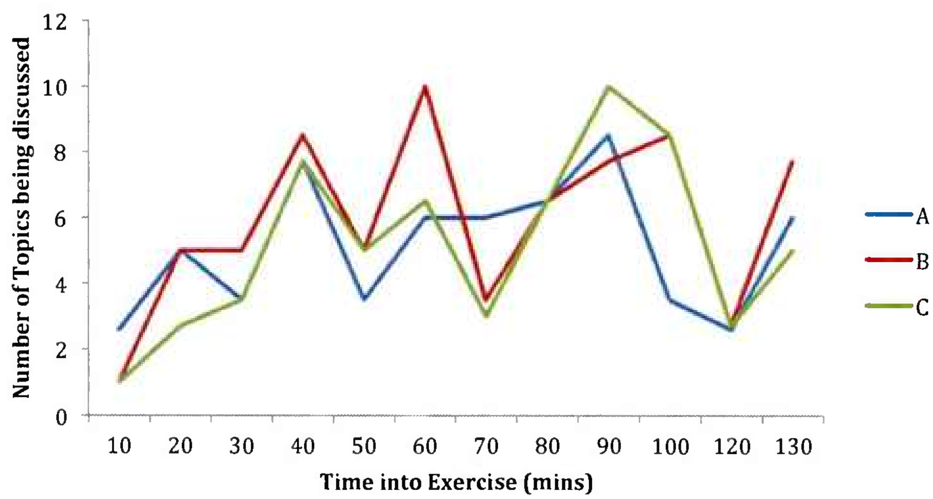
[6] A complete pack of materials can be obtained from the lead author on request.

RESULTS AND ANALYSIS

In these data, counts of activity (related to number of topics discussed, number of actions performed etc. during the sampling window) are presented in the form of graphs to provide a convenient means of comparing groups. Deeper analysis is provided in the form of extracts from qualitative analysis of specific events in the exercises.

Activity Sampling Results

The evidence cards contained information which can be classified in terms of Suspect, Date / Time, Locations (Exeter, Roskoff, Leeds, Exmouth), Vehicle (yacht, van) Action (payment, social activity, business, crime). This classification defines the set of topics that groups discussed. In the activity sampling, each mention of a topic was counted in the sampling period. Thus, if the group said 'Condiere owns the yacht called Sunny Jim', we would count 1 for 'suspect – Condiere' and 1 for 'vehicle – yacht'.



Figue 2 : Number of topics discussed over time

Figure 2 plots the number of topics mentioned at each sampling period in version two. Overall, the average number of topics is consistent across the groups, i.e., 2A = 6 (± 3), 2B = 5 (± 4) and 2C = 5 (± 3), which are similar to the results of three of the groups in version one.

Processes of Sense-making

From the activity sampling data, it is possible that the groups tended to alternate between broad (several topics) and narrow (few topics), which suggests movement from linking to the development of rules. To explore these forms of interpretation, the following section presents extracts of discussion between participants. The discussions are verbatim records of participants' statements. Each extract is identified by group {A,B,C} and speaker {a,b,c,d,e}.

SEEKING A FRAME

Group B (5 participants) began by discussing the Angel warehouse and then split into two sub-groups. One subgroup searched for more information about the warehouse, while the other subgroup (consisting of Participants Bb and Be) developed a social network diagram on the whiteboard. Thus, Group B framed the task as a social network problem. They identified a possible American connection (presumably in terms of the purchase of the marina by a US company in a deal brokered by Grosby). Participant Be adds "American connection?" to the bottom of the social network diagram. No new representations were created, and the idea of a timeline was raised but then dismissed.

DEVELOPING A FRAME

Group C argued between arresting Sarti and arresting Pasquidini. In particular, group 2C discuss the role of Pasquidini and present inferences that feature him as a shadowy figure (Cb "the anonymous Frenchman") who Ca notes is "connected…He has links", and Cd observes "He's been calling all over the joint". Against this proposal Cb is concerned "All we have is the phone records" which raises the issue of what would constitute evidence to support the arrest of an individual in this exercise. This discussion leads to the development of a rule proposed by Ca "We need to take out the source in France to bring down the whole network."

The initial link diagram drawn by group C resulted in Pasquidini being in the centre of the network (this was by coincidence as names were added to the diagram, not by intent or design by the person drawing the network). As the discussion progressed, the group became convinced that Sarti was the lynchpin of this activity. Consequently, the diagram was erased and a new diagram drawn starting with Sarti in the centre.

Group C shifted their attention between different suspects (Condiere, Sarti, Calabrese, Ricord, Pico and Grosby) and attempted to define associations between these suspects. Thus, Ce notices (from reviewing financial statements that Sarti is receiving unusual payments, Cd (reviewing telephone records) notes that Sarti is "linked to everyone", and Cc (from the addresses on suspect cards) observes that Sarti "lives next door to Pico" from which Cb infers that Pico "obviously knows her". Having emphasised these connections, Cb still concludes that "there's not enough evidence" and that "Sarti can't be the driver because Calabrese was arrested. He took the van to Exeter." They also consider Pico who, as Cc points out, "They paid his bail so they don't want him talking". This illustrates how, even in the space of the 5 minutes over which these exchanges took place, the teams work with multiple frames but rarely develop these into concrete hypotheses.

COMPARING, ELABORATING AND QUESTIONING FRAMES

In this exercise, the groups would regularly (i.e., every 20 minutes or so) collect around the representation that they were creating and run through their analysis. At one level, this could be seen as rehearsal of their final presentation. At a deeper level, we propose that this run through provides an opportunity to elaborate and question the story that best explains the analysis. In other words, the groups applied effort to elaborating and questioning the frames they were using (i.e., 'conflict and corroboration' and 'hypothesis exploration' (Elm et al., 2005), and their rehearsals could be seen as ways of testing the narrative of their analysis.

Comparison of Solutions

In terms of solution, all three groups identified Sarti (3/3) as a prime suspect (because she was central to so much of the exercise) and all three teams named Pasquidini (3/3). Two of the groups proposed that Pasquidini should be arrested in France by Interpol, suspecting his involvement in supplying the drugs and loading them on to the yacht. The groups also discussed Pico / Cobo (2/3), as a possible driver of the van, and mentioned Calabrese (2/3) in support of this proposal; Calabrese was Ricord's chauffeur and had been arrested driving the van, and now Pico / Cobo was Ricord's chauffeur so looked suspicious. All three groups used the Harbour Logs (3/3) and Accounts (3/3) to provide evidence of who was linked to whom and when events occurred, with the phone records (2/3) supplementing the links.

DISCUSSION

The design of the Exercise had assumed that people would identify the modus operandi (M.O.) of the gang and then look for information as to when this M.O. was likely to be applied. However, while there are instances where the groups described the M.O., this did not seem to be the primary focus on their analysis. Rather, data were combined into sets, or frames, and these frames explained or represented.

Using Frames

Frames begin in a sketchy (minimal) manner, either through the linking of data in representations or through the linking of concepts in response to questions. This might be a function of the nature of the evidence provided for the exercise, with all groups beginning their processing with the sorting of sheets into piles. An interesting point to note here is when, or if, the piles of evidence became 'frames'.

The groups not only sought links between sampled evidence, but also developed their hypotheses through testing them. Thus, group C not only raised hypotheses about Pasquidini, Sarti and Calabrese but also challenged these hypotheses in their discussions. This suggests that these groups were not only engaging in the 'down-collecting of material but also in 'conflict and corroboration'. Further, the observation that all the groups would rehearse their presentation at intervals during the Exercise suggests that they recognised the value of or 'hypothesis exploration' as a core part of their analytical work.

In terms of the Data/ Frame concept of sense-making, this paper offers some insight into the dynamics of the process of sense-making in teams. The observational data suggest that teams prefer to work with a small number of pieces of evidence (i.e., a mean of 5 pieces, irrespective of experience). Further, the groups tend to move between broad and shallow and deep and narrow search (focusing on specific frames, but, generally working with more than one frame at any one time). This suggests that traversal of the Data / Frame cycle is more effective for the experienced analysts, who were able to switch frames. Previous work had shown that the less experienced analysts might have been either unable to generate an appropriate frame (being swamped by data) or who might have stuck with a particular frame even when it is not appropriate to the current set of data (Baber et al., 2013). This implies that differences in sense-making of experienced and inexperienced analysts are not simply a matter of knowledge but also relate to the manner in which evidence is selected and processed, and hypotheses and frames employed.

The extracts of team discussions suggest that, even when teams focus on a frame, their attention is drawn to other data and the analysis moves between several frames in short succession. This suggests that traversal of the Data / Frame model is faster than one might expect. In other words, in this Exercise, teams seem to move through the Data / Frame cycle quickly, with consideration of several frames, rather than taking a single frame and processing this. This implies that the Exercise resulted in an abductive approach to reasoning, in which the data were explored and resulting frames considered, rather than a deductive approach in which hypotheses are raised and tested. While we would not claim that this represents all forms of Intelligence Analysis, it is interesting that this cyclical approach is very different from the linear approaches implied by Pirolli and Card (2006) or Heuer (1999). What we observed in this study was that, while people operate using 'competing hypotheses', these tended to be articulated as loose, imprecise statements rather than as objectively grounded comparisons. While this is likely to be an artefact of the study (and is not meant to imply than Intelligence Analysis does not or should not seek objective grounding of hypotheses), it does suggest that searching for problem opportunities is as much an art as a science.

Using Representations

Representations are a way of externalizing a frame in which sets of data can be combined. The representations either focused on the grouping of people (through link or social network diagrams) or events (through timelines), or a combination of these. Initially these were a means of grouping data. However, as the Exercise developed, the representations became the focus of the final presentation. This meant that, rather than creating representations to serve as aide memoire for their own discussions (as inexperienced groups did in Baber et al., 2013), the groups were creating representations for an audience, i.e., their Commanding Officer to whom they would give their presentation. This suggests that the activity was primarily a hypothesis creation activity in which relations between topics and resulting inferences could be used to create hypotheses for further investigation.

As de Vries and Masclet (2013) point out, collaboration is very often based on minimal representations. In this Exercise, representations created in collaborative activity are not merely diagrams showing data; rather, they are records of the discussion and thought-processes of the groups. This means that, in order to understand the content of the representations, it is often necessary for someone from outside the group to have an explanation of the assumptions, ideas and background knowledge that inform these representations. In other words, the role of representations is often to capture 'local' discussion rather than to create a more 'global' view. In order to develop from this local to global view of the information, it is important to have some notion of audience. Intelligence Analysts often talk of 'product' as the output of their activity. What we note here is that 'product' can look much the same whether it is produced by experienced or inexperienced analysts, and the primary difference is not the 'product' per se so much as the understanding of who will use that product, how they will interpret the product, and what aspects of the product they will find convincing. This distinction supports the advice offered by Heuer (1999).

Conclusions

The study in this paper supports the observation that Intelligence Analysis is not a linear, orderly process (see also Elm et al., 2005; Kang and Stasko, 2011; Roth et al., 2010). Even with so simple a set of evidence, we could observe behaviour which was parallel (with several group members working on different lines of enquiry), disjointed (with group members pursuing contradictory 'frames', e.g., arrest Sarti or Pico, or Sarti or Pasquidini), and recursive (with groups dismissing a frame and then reintroducing it, e.g., dismissing the abandoned car and then considering that it was used as the drug transport vehicle). This suggests that such behaviour is likely to be a characteristic of this type of activity. From this, it is apparent that the activity is primarily one in which small sets of data are combined and explained.

From the use of representations, it is apparent that experience dictates the manner in which people construct, use and share representations. This suggests that the design of "sense-making support systems" (Weick and Meader, 1993) should not focus simply on ways to support the construction of diagrams and other forms of representation, but also needs to consider the manner in which these representations are to be used. For example, tools which support the collation of information to help identify links between pieces of information might help with 'down-collection' of data but does not provide support for 'conflict and corroboration' or for ' hypothesis exploration'.

REFERENCES

Attfield, S. and Blandford, A. (2010) Discovery-led refinement in e-discovery investigations: sensemaking, cognitive ergonomics and system design, *Artificial Intelligence and Law, 18,* 387-412.

Baber, C., Attfield, S., Wong, W. and Rooney, C. (2013) Exploring sensemaking through an Intelligence Analysis exercise, *NDM: 11^th^ International Conference on Naturalistic Decision Making,* Marseilles, France.

Baber, C. and Butler, M., 2012, Expertise in Crime Scene Examination: comparing search strategies of expert and novice Crime Scene Examiners in simulated crime scenes, *Human Factors, 54,* 413-424.

Cabitza, F., Colombo, G. and Simone, C. (2013) Leveraging underspecification in knowledge artifacts to foster collaborative activities in professional communities, *International Journal of Human-Computer Studies, 71,* 24-45

Chase, W.G. and Simon, H.A. (1973) Perception in chess, *Cognitive Psychology, 1,* 55-81.

Cook, M.B. and Smallman, H.S. (2008) The human factors of confirmation bias in intelligence analysis: decision support from the graphical evidence landscape, *Human Factors, 50,* 745-754.

Dervin, B. (2003) A sense-making methodology primer: what is methodological about sense-making, *Meeting of the International Communication Association*, San Diego, CA.

de Vries, E. and Masclet, C. (2013) A framework for the study of external representations in collaborative design settings, *International Journal of Human-Computer Studies, 71*, 46-58

Didierjean, A. and Fernand, G. (2008) Sherlock Holmes – an expert's view of expertise, *British Journal of Psychology, 99,* 109-125.

Elm, W., Potter, S., Tittle, J., Woods, D., Grossman, J. and Patterson, E. (2005) Finding decision support requirements for effective intelligence analysis tools, *Proceedings of the 49th Annual Meeting of the Human Factors and Ergonomics Society*, Santa Monica, CA: HFES, 297-301.

Heuer, R. (1999) *Psychology of Intelligence Analysis*, Washington, DC: Central Intelligence Agency, Center for the Study of Intelligence.

Heuer, R. and Pherson, R.H. (2010) *Structured Analytic Techniques for Intelligence Analysis*, CQ Press.

Kang, Y. A., Görg, C. and Stasko, J. (2009) Evaluating visual analytics systems for investigative analysis: Deriving design principles from a case study, *IEEE Symposium on Visual Analytics Science and Technology (VAST 2009)*, New York: IEEE, 139-146.

Kang, Y.A. and Stasko, J. (2011) Characterizing the intelligence analysis process: Informing visual analytics design through a longitudinal field study, *IEEE Conference on Visual Analytics Science and Technology (VAST 2011)*, New York: IEEE.

Klein, G., Moon, B.M. and Hoffman, R.R. (2006) Making Sense of Sense-making 1: Alternative Perspectives, *IEEE Intelligent Systems, 21*, 70-73.

Klein, G., Moon, B.M. and Hoffman, R.R. (2006) Making sense of sense 2: A macrocognitive model, *IEEE Intelligent Systems, 21,* 88-92.

NATO (2008) *AAP 6 Terms and Definitions*

Pirolli, P. and Card, S. (2005) The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis, *Proceedings of the International*

Conference on Intelligence Analysis, McLean, VA: Office of the Assistant Director of Central Intelligence for Analysis and Production.

Roberts, N. C. (2011) Tracking and disrupting dark networks: Challenges of data collection and analysis, *Information Systems Frontiers, 13,* 5-19.

Roth, E.M., Pfautz, J.D., Mahoney, S.M., Powell, G.M., Carlson, E.C., Guarino, S.L., Fichtl, T.C. and Potter, S.S. (2010) Framing and contextualizing information requests: problem formulation as part of the intelligence analysis process, *Journal of Cognitive Engineering and Decision Making, 4,* 210-230.

Stasko, J., Görg, C. and Liu, Z. (2008) Jigsaw: supporting investigative analysis through interactive visualization, *Information Visualization, 7,* 118-132.

Tecuci, G., Schum, D., Boicu, M., Marcu, D. and Hamilton, B. (2010) Intelligence analysis as agent-assisted discovery of evidence, hypotheses and arguments, *Advances in Intelligent Decision Technologies*, Berlin: Springer, 1-10.

Trent, S.A., Patterson, E.S. and Woods, D.D. (2007) Challenges for cognition in intelligence analysis, *Journal of Cognitive Engineering and Decision Making, 1,* 75-97.

Treverton, G. F. and Gabbard, C. B. (2008) *Assessing the Tradecraft of Intelligence Analysis,* Santa Monica: Rand.

Treverton, G.F. (1994) Estimating Beyond the Cold War, *Defense Intelligence Journal, 3,*

Weick, K.E. (1995) *Sensemaking in Organisations,* London: Sage.

Weick, K.E., Meader, D.K. (1993) Sense-making and group support systems. In: Jessup, L., Valacich, J. (Eds.), *Group Support Systems: New Perspectives*, Macmillan Publishing Company, New York, NY, USA, pp. 230–252.

Wu, A., Convertino, G., Ganoe, C., Carroll, J. M., & Zhang, X. L. (2013) Supporting collaborative sense-making in emergency management through geo-visualization, *International Journal of Human-Computer Studies, 71*, 4-23.

Appendix

1. The Solution is to arrest Pasquidini (the passenger on the yacht, as outlined in the M.O. above). The puzzle is to place Pasquidini in Exmouth and to see him as the van driver. There is no direct evidence to this effect (which is why the Exercise is challenging). However, the combination of M.O. and evidence from today and yesterday should help the groups narrow down their set of suspects and realise that the yacht's passenger is the van driver and that Pasquidini possibly travelled from France.

2. There are three people who are dubious but who have insufficient information to justify arrests:
    a. Muriel Grosby has been involved in the deal to buy the marina and has a wide range of highly suspicious transactions in her business accounts. She also owns the mini-cab firm which is dealing the drugs. On the other hand, she is involved in charitable events with the Marina and with making donations to it. The accounts and client list of Ricord Accountancy Services link many of the characters together suspiciously – but not in sufficient detail to clearly indicate nefarious activity.
    b. Martina Sarti hires the vans which are used for transporting the drugs – but it is likely that she hires vans on a regular basis for people coming into the marina and not specifically for the smuggling operation. She has received money from Grosby but it is not obvious why this is suspicious, given their relationship with the marina.
    c. A petty criminal (Cobo a.k.a Pico) who comes from Leeds, is living next to Martina Sarti (although it is likely that she is spending most of her time in Exmouth with David, the marina manager), and is being paid as a chauffeur by the accountant Ricord.

In order to arrive at the solution, one approach would begin with the arrest of Calabrese, who (as pointed out in the briefing was sentenced on 14th June 2012). The newspaper article detailing Calabrese's sentencing notes that he was arrested in November 2011. Two statements dating from November 2011 (one from Calabrese and one from Bocognani, the former manager of the Marina) suggest that the gang's M.O. is to ship drugs from Roskoff on a yacht skippered by Perrin, to arrive at Exmouth in the early hours and for the drugs to be moved by van to the Angel Warehouse in Leeds. A record of van hire shows that Sarti hired a van in early November

2011. A review of other van hire logs shows that Sarti hired a van in August 2012 and hired a van yesterday (9th September 2012). The Marina log shows that the only yacht due in today is the 'Sunny Jim', owned by Condiere. The other evidence that corresponds to 'today' is the phone logs of Pasquidini, who calls Condiere, Perrin, Angelleti, Munoz and Sarti. Pasquidini's 'suspect card' shows that he lives in Roskoff.

Narrative Visualisation: designing User Interfaces to support Sense-Making through Visual Analytics

Adam DUNCAN[a], Chris BABER[a], and Natan MORAR[a]
[a]*School of Electronic, Electrical and Systems Engineering, University of Birmingham*

ABSTRACT
In this paper we present a story-tiling visualization technique, and a study comparing it with concept mapping. With this research we are exploring how visual narratives might aid in the processing and organization of information during a sense-making task. We suggest based on our findings that the creation of a linear, storytelling diagram may assist analysts in identifying where information does not fit within a coherent narrative, but that a more open-ended diagramming technique allows for multiple strands of information to be incorporated. Our aim in this research is not to challenge or replace concept mapping, but to suggest a possible alternative designed around storytelling for these kinds of sense-making tasks.

KEYWORDS
Sense-making; User Interface Design ; Narrative.

INTRODUCTION

Visual Analytics solutions attempt to combine both human and automated data analysis into the best of both worlds. The visual analytics process (Keim et al., 2010) starts with the (automated) transformation of data which can then be mapped to graphical properties of visual representations ready for exploration and analysis. For automated analysis, data mining methods are used. In the case of visual analysis, the analyst can manipulate visualisations in order to interact with the automated process. This interaction could include database queries or algorithm adjustment. This process is illustrated by figure 1.

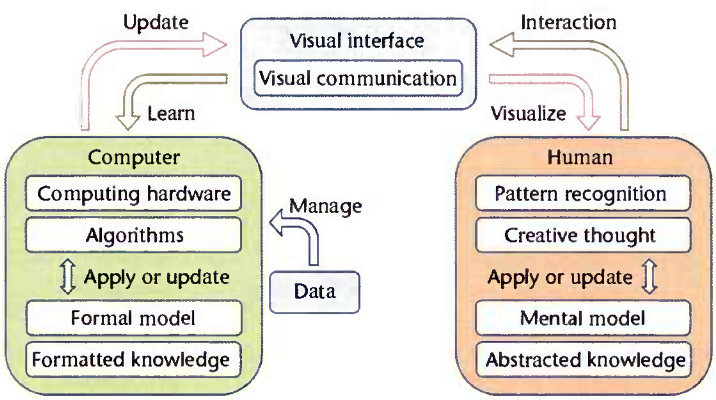


Figure 1 : The Visual Analytics Process (from Mueller et al., 2011)

The ultimate goal of Visual Analytics is to generate knowledge and insight from the data (Thomas and Cook, 2005). As Mueller et al. (2011) see this this activity as a form of iterative learning through which the user constructs a model of the analytic problem and reviews the data accordingly. From the perspective of Naturalistic Decision Making, the task of the analyst is to make sense of the data and its analysis which can be considered in terms of the Data/ Frame model of sense-making (Klein et al., 2006a,b). In this paper, we explore the design of Visual Analytics, in terms of the Data / Frame model and offer a novel design for visualising narrative as an additional medium to support analysis. The remainder of the paper has three sections: in the next section, we present a mapping between Data / Frame model of sense-making and tasks in Visual Analytics, following this we argue that the visualisation of narrative has, to date, received less attention than conventional approaches to visualisation, which leads to a study which compare story-tiles (our technique for visualising narrative) with concept maps.

User tasks

There are are variety of approaches to the description of user activity with Visual Analytics. Shneiderman (1996) proposed a task x data-type taxonomy of seven tasks {overview, zoom, filter, details-on-demand, relate, history, extract} and seven data types {1-dimensional, 2-dimensional, 3-dimensional, temporal, multi-dimensional, tree and network}. In their description, Yi et al. (2008) suggest that users generate insight by following an iterative process in which they seek to overview the data, then look for patterns in the data, adjusting the data in order to match a set of expectations (or mental model) that they bring to their analysis (or which is implied by the pattern that they have detected).

Visualisation

Visual analytics tools incorporate a wide array of views for dealing with complex data and sensemaking tasks. Faisal et al. (2009) classify six common representational types created for and relied upon in the sense-making process: spatial, argumentational, faceted, hierarchical, sequential, and networked. Networked and hierarchical (the example given being a treemap) representations are again noted amongst these standard visualisation techniques. A review of fifteen commercial visual analytics software, conducted by Zhang et al. (2012), supported the following visualisation techniques: bar, line and pie charts, histograms, scatterplots, heatmaps, and map overlays. Other graphical representations included parallel coordinate plots, scatterplot matrices, treemaps and network graphs.

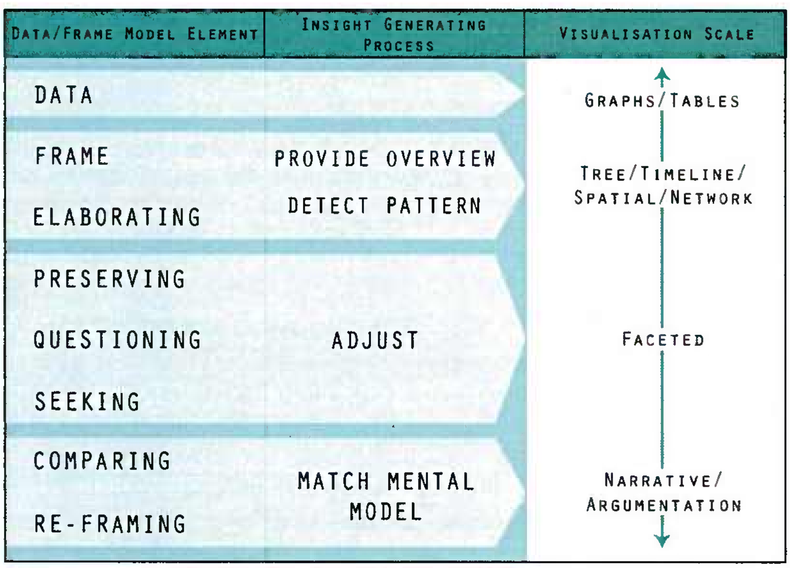


Figure 2 : Mapping user tasks and visualisation to the Data / Frame model of sense-making

Figure 2 shows our attempt to map the stages of the data/frame model with the processes of insight generation defined by Yi et al. (2008), and with potential visualisation solutions. The visualisations on our scale are derived from Faisal et al. (2009), Segel and Heer (2011) and from our own work with visual representation. We believe that graphs and tables provide a straightforward visualisation representation of data, but don't allow for more complex frames that are required when interpreting qualitative information. Other visualisation types allow for more elaboration and complex framing (e.g. temporally or spatially), this in turn can help with the detection of patterns and relationships within complex data. On the other end of the scale, we feel that there are visual representations that can help analysts or researchers create a mental model that describes particular information or events in sense-making. These could take the form of narrative (which may help to describe what is thought to have happened or a sequence of events), or argumentation.

NARRATIVE VISUALISATION AS THE MISSING LINK

Visualizing narratives and using visualization to support storytelling is not new. Segel and Heer (2010) present seven genres of narrative visualization, e.g., magazine style, annotated chart, partitioned poster, flow chart, comic strip, slide show and video. We suggest that, while the concept of incorporating narrative visualization techniques, such as comics and storyboards, into applications for visual analysis, is not entirely novel (ap Cenydd et al., 2011; Jin and Szekely, 2010 ; Hossain et al., 2012), it is of growing interest and there is still significant research to be done in understanding how we can work with storytelling approaches in this field. In other words, telling the story with the data is a skill that the analyst brings to the presentation of the analysis rather than a fundamental feature of the visualisation itself.

We develop analyst-constructed narrative visualisations, as opposed to the analyst-driven or reader-driven visualisations noted by Segel and Heer (2010). Our diagramming tool produces Story-tile visualisations (see figure 3). Story information (e.g. places, characters, actions and other information) is encoded as icons and graphical objects within a series of 'scenes.' Each scene is a tile in the representation. A sequence of tiles describe a sequence of events within a narrative. The approach is based on a comics, or perhaps more accurately a storyboarding metaphor. Segel and Heer's (2010) genres of narrative visualisation list sequential methods such as comics and slideshows, a storyboarding metaphor has also been explored in relation to visual analytics but it focused on different problems, types of data and arrived at an alternative end result (ap Cenydd et al., 2011).

The process of creating story-tiling begins with the extraction of information from a source, in the case of this study through manual selection from a source document (although other possibilities could include named-entity recognition partially automating this process). An icon is created (see figure 3) from the extracted information that represents an entity, this entity then links to its parent document, can be manipulated and can be 'opened up' to reveal metadata attached to it. For this study we limit them capabilities to just the representation of the entity. We did this to reduce the number of potential variables and to keep the two tools similar in terms of capability and to reduce the complexity of learning the interfaces. The user then constructs 'scenes' that describe who was in a location, with whom, what they did there and when. The tiles that contain the scenes are generated using buttons created dynamically by the interface (the buttons appear where a tile can be added to sequence).

Figure 3 : Creating a tile from source document

COMPARING NARRATIVE VISUALISATION WITH ARGUMENTATION

Concept Mapping

Concept maps are a flexible, diagramming technique that have already been evaluated in a number of areas (Moon et al., 2011). It should be noted that the visualization tool we developed for concept mapping has some differences to traditional concept mapping, specifically the concepts created are selected from the documents in the dataset (as opposed to the user having complete freedom in defining concepts). We wanted to examine what information is treated as important and gathered by participants, so we felt that we needed to identify where that information had been extracted from. See figure 4 for a simple example of concept map produced by a user working with our software Derbentseva and Mandel (2011) developed a concept map knowledge model to support a technical report summarizing an exploratory interview study with a sample of managers from Canadian intelligence organizations. The purposes behind the development of their Concept Map were the same as those suggested by Heuer (1999): to help organize thinking and achieve an understanding of key concepts, and to facilitate communication of complex relationships.

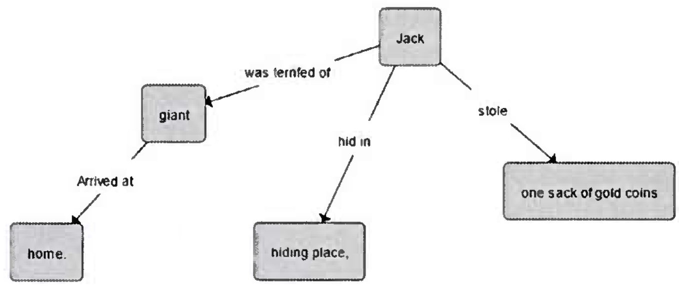


Figure 4 : Simple concept map

Method

This study involved 14 participants (4 female and 10 male; age range 22 to 58 years) who were unfamiliar with the software tools or the information used in the experiment. Participants navigated to a specific URL to log on to the experiment. The experiment was approved by the University of Birmingham Ethics Committee. The initial screen explained the purpose of the experiment and asked participants to indicate their gender and age range and to confirm that they accepted that we could use their data for our analysis.

The experiment was run as a repeated measures design. Although we accept that the simplicity of the task was likely to support learning effects, we were interested in subjective comparison of the media. Participants were then assigned to one of two conditions: concept map or story-tiles. In each condition, participants were required to search a set of documents and select information which they used to construct visualisations which best fitted their interpretation of a fictitious terror event (taken from the VAST 2011 Challenge). In order to make the experiment tractable, we present a set of 10 documents (culled from the 4400 originally presented in the VAST challenge). The reason for this limited selection is that we were not employing any form of automated data analysis and data reduction, so each document had to be read and reviewed by hand. Our set of documents comprised four relevant documents (containing information pertinent to the task), two false leads (containing information about terrorist activity but not relevant to the main task) and four noise documents (containing irrelevant information).

The concept mapping based tool was not quite traditional in the sense that existing tools (such as CmapTools) are. Our tool only allows concepts to contain textual information selected from the documents. We took this approach to see exactly where participants selected information from and then how they structured it into a visual form, as well as to keep the two visualization processes similar. It would have become more difficult to analyze where information was derived from if participants had been allowed to input their own concepts. The tool worked by allowing the user to select a keyword or a group of words from the document currently being viewed, these words would then be added into their visualization in the form of a concept box. Users could then position this (or any other existing) boxes and link them together by clicking on one and then another. Once linked users could label the relationship between two concepts. Concepts could be deleted from a visualization by clicking on an undesired box and then clicking the delete button, deletions were recorded by the application. The second visualization allowed users to produce a story-tiling visualization (see Story-tiling Design section for more detailed information on the design of story-tiling). In this instance participants could extract information to categories as an entity by selecting individual words or groups of words from the current document in the same way as the concept mapping based tool. However, after a selection had been made it could then be categorized by the participant as a particular entity type using a drop-down menu. This entity type is then be added to the current story tile (as an icon) which collates the various entities related to or involved in a particular event. The icons can then be moved around the tile, once the participant is satisfied with the tile configuration they can then position the tile in the story sequence (a collection of tiles that tell the desired story).

Data collection and analysis

Data for the experiment was collected by the application as the participants progressed with their task. Relevant information was stored on the server. The visualizations that were created by participants were recorded with all of the information necessary to recreate them, and additionally information about which document a particular piece of information had been selected from. The analytic 'history' of the participant was recorded as a string, it took the format of: document viewed, selections and deletions made from that document, the time at this stage since the application start, next document viewed and so on.

RESULTS

There was no differences in terms of items in the final representation (U=22.5, n.s.) ; both groups tended to have similar numbers of items, i.e., 22 items in the Story-tiles and 21 items in Concept Maps. Participants using Story-tiles, however, tended to delete more items than those using concept maps (i.e., 13 items deleted during the course of the trial when using Story-tiles compared with 6 deletions when using Concept Maps), although this was no significant (U = 13.5, n.s.). A higher number of items selected by participants using Story-tiles (35 items) compared with those using Concept Maps (28 items), although, again, this was no significant (U =17, n.s.).

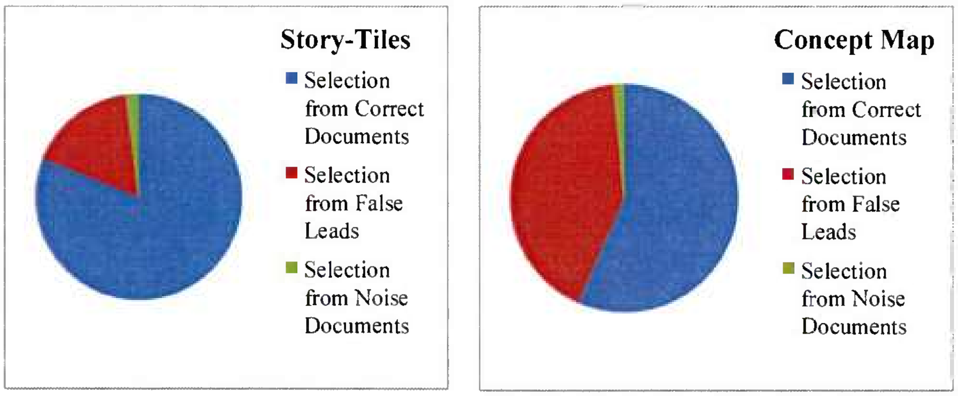


Figure 5: Relative distribution of information sources (as a % of items in the final visualisation)

As figure 5 shows, both groups were able to ignore the 'noise' documents in their selection of material. However, participants using Concept Maps were more likely to select material from the 'false leads' than those using the Story-tiles. Comparison of the groups showed that participants using the Story-tile had significant higher proportion of selection from 'correct' documents than those using Concept Map (U=6, $p<0.05$) and participants using Concept Maps had significantly higher selection from 'false leads' than those using Story-tiles (U=3, $p<0.05$).

Figure 6 shows a bubble matrix, which shows the mean time participants spent working with each document, and how much information they gathered from a particular document; the larger the bubble, the more time was spent on a document, and the darker the bubble, the more information was selected from the document. The background colour of each cell indicates whether the document was noise (white), false lead (orange) or correct (green). Participants spent most time working with the first document (a false lead) but subsequently devoted more time with relevant documents than with false leads or noise. If we recall that false leads were more commonly used in the Concept Maps then more filtering seemed to take place in the story-tiling. One suggestion is that participants using Story-tiles collected more items but also put more effort into editing the visualisation (in terms of removing some of the selected items).

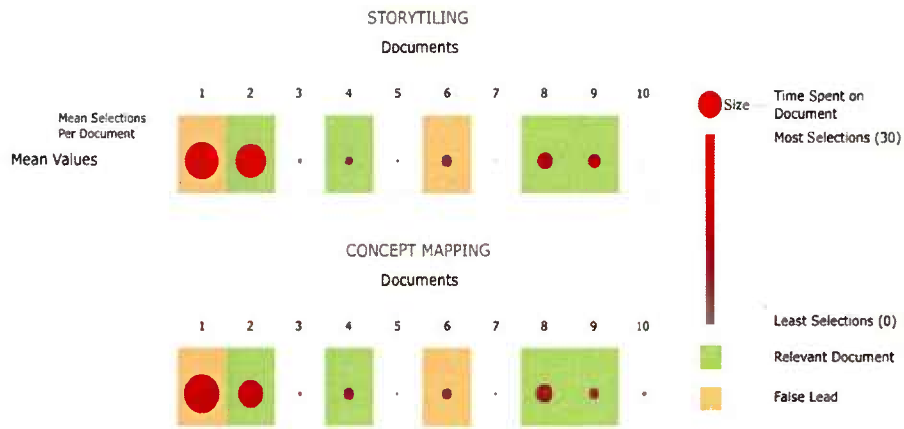


Figure 6: Bubble Matrix comparing media

Observations and Conclusions

The results from this study suggest a difference in construction strategy between the two media. Using the story-tiles, participants were building and editing their stories throughout the task with more emphasis on collecting and deleting material. In contrast, during concept mapping participants collected less information but deleted less. This difference, in deletions, is further highlighted by the higher 'false positive' rate in the Concept Maps. We propose that, when using story-tiles, participants were constructing linear narratives and selecting information to fit the narrative that they are building (making it difficult to incorporate information which does not fit coherently). When using Concept Maps, the structure of the information was seen as more flexible and participants incorporated several strands of information, and were reluctant to lose information. In story-tiles, participants work within the constraints of the structure created by an understanding of the linear flow of a story, whereas Concept Maps were used to create a structure on a more ad hoc basis. It should be noted at this juncture that, despite instruction in the correct use of Concept Maps (in terms of defining logical structure when reading the diagram) participants tended to use the tool to create a spatial rather than logical arrangement. This means that, rather than being read as a critique of Concept Maps, the study shows how participants were misusing the tool.

DISCUSSION

The aim of this study was consider how visualisation structure can affect analysis processes and strategies. Where data are quantitative then there are tools and techniques which allow analysts to represent and analyze those data, e.g., in the form of social network graphs. However, where those data are qualitative, e.g., witness testimony, news reports etc., it can be difficult to either produce compelling visualizations or accounts. We explored the potential benefits of storytelling visualizations and have introduced our story-tiling technique to this end.

Our study has highlighted potentially interesting effects in the capture of information (both in terms of quantity, and source used). This analysis suggests that participants approached the selection of information in different ways depending on the medium used to represent the data. We suggest that it may be more difficult to fit irrelevant information into a linear, structured medium because the narrative will lose coherence. In a diagramming technique, like Concept mapping, different strands of information (including possibly incorrect or irrelevant ones) are more easily incorporated into a whole picture.

Participants commented that the ability to define their own relationships was useful within Concept Maps, with one participant saying "the story-tiles' categories were useful, while in the text-based diagrams I liked being able to connect items and label the connections" and another commenting "the ability to annotate the story board would have been really useful." This suggests that participants would like to have more input to clarify relationships within a story-tiling environment. As noted previously these capabilities were not included in the tested story-tiling tool for fear of complicating both the evaluation and increasing the learning curve. Our intention is to continue to research and improve the story-tiling visualization technique; introducing the ability to add users defined content into tiles.

In the data-frame model of sense-making, frames are applied to data in order to help make sense of it. These frames are rejected when they are no longer considered helpful, and the process of reframing is an iterative one. Story-tiling participants continued to collect information evenly throughout the activity, and filtered out more of these selections. Perhaps story-tiling may cause the re-evaluation of frames throughout the sense-making activity, as new data points cannot be incorporated into alternative threads within the representation. This suggests an interesting route for further research into the implications of complex, branching, visual narratives versus linear imagery which tells a less ambiguous tale. The results have also raised questions for further study regarding whether or not visualization rules could be more rigidly enforced by a system, whether that would impair usefulness and what level of freedom is required to effectively assist users in understanding concepts from a document collection.

REFERENCES

ap Cenydd, L., Walker, R., Pop, S., Miles, H., Hughes, C., Teahan, W., & Roberts, J. C. (2011, October). epSpread-Storyboarding for visual analytics. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*(pp. 311-312). IEEE.

Derbentseva, N. and Mandel, D.R. Using Concept Maps to Improve the Practice and Organization of Intelligence in Canada. In B.M. Moon, R.R. Hoffman, J.D. Novak and A.J. Cañas, eds., *Applied concept mapping: theory, techniques, and case studies in the business applications of Novakian concept mapping.* CRC Press, Boca Raton, FL, 2011, 109–130.

De Vries, E. and Masclet, C. A framework for the study of external representations in collaborative design settings. *International Journal of Human-Computer Studies 71*, 1 (2013), 46–58.

Faisal, S., Attfield, S., and Blandford, A. A Classification of Sensemaking Representations. (2009).

Heuer, R.J. (1999) *Psychology of Intelligence Analysis*, Washington: Center for Study of Intelligence.

Heuer Jr, R.J. and Pherson, R.H. (2010) *Structured Analytic Techniques for Intelligence Analysis.* Washington, D.C.: CQ Press.

Hossain, M.S., Butler, P., Boedihardjo, A.P., and Ramakrishnan, N. Storytelling in entity networks to support intelligence analysts. *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM (2012), 1375–1383.

Jin, J. and Szekely, P. Interactive querying of temporal data using a comic strip metaphor. *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, (2010), 163–170.

Keim, D.A., Kohlhammer, J., Mansmann, F., May, T., Wanner, F. (2010) Visual Analytics, in: Keim, D., Kohlhammer, J., Ellis, G., Mansmann, F. (Eds.), Mastering the Information Age: Solving Problems with Visual Analytics. Eurographics Association, Goslar, Germany, 7–16

Klein, G., Moon, B.M. and Hoffman, R.R. (2006a) Making Sense of Sense-making 1: Alternative Perspectives, *IEEE Intelligent Systems, 21*, 70-73.

Klein, G., Moon, B.M. and Hoffman, R.R. (2006b) Making sense of sense 2: A macrocognitive model, *IEEE Intelligent Systems, 21,* 88-92.

Moon, B.M., Hoffman, R.R., Novak, J., and Cañas, A., eds. *Applied concept mapping: theory, techniques, and case studies in the business applications of Novakian concept mapping.* CRC Press, Boca Raton, FL, 2011.

Mueller, K., Garg, S., Nam, J.E., Berg, T., McDonnell, K.T., 2011. Can Computers Master the Art of Communication?: A Focus on Visual Analytics. Comput. Graph. Appl. IEEE 31, 14–21.

Segel, E. and Heer, J. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (2010), 1139–1148.

Shneiderman, B., 1996. The eyes have it: a task by data type taxonomy for information visualizations. Presented at the IEEE Symposium on Visual Languages,

Thomas, J. and Cook (2005) *Illuminating the Path: The Research and Development Agenda for Visual Analytics*, United States of America: National Visualization and Analytics Center.

Yi, J.S., Kang, Y. and Jacko, J.A. (2008) Understanding and Characterizing Insights: How Do People Gain Insights Using Information Visualization, *Proc. 2008 Conf. beyond Time and Errors (BELIV 08)*, 1–6.

Zhang, J. and Norman, D.A. Representations in Distributed Cognitive Tasks. *Cognitive Science 18*, 1 (1994), 87–122.

Zhang, L., Stoffel, A., Behrisch, M., Mittelstadt, S., Schreck, T., Pompl, R., Weber, S., Last, H., Keim, D., 2012. Visual analytics for the big data era—A comparative review of state-of-the-art commercial systems, in: Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on. IEEE, pp. 173–182.

Naturalistic Decision Making, Risk Management, and Leadership

MAJ Bridgette Bell[a] and COL James Ness[a]
[a]*Engineering Psychology Program, Department of Behavioral Sciences and Leadership, United States Military Academy, West Point, NY*

ABSTRACT
Risk guides the way groups work together, the way organizations learn, and how much trust individuals have in one another. Organizations rely on human interaction to accomplish intricate missions and solve complex problems by employing risk management processes. We recommend further investigation into risk management within the naturalistic decision making framework to determine how leaders accomplish missions through assessment of work processes and personnel. As Army leaders aim to seek assistance for their soldiers, they are constantly assessing the value of available resources and determining risks at different levels. Further dissecting risk management into the following constructs will help us address more effective leadership decision-making : fear of unknown knowledge, assessment of failure, efficiency in thinking, and productive mission accomplishment.

KEYWORDS
Naturalistic decision making, risk management, leadership, suicide prevention, Army

INTRODUCTION
Human Systems Integration encompasses an approach to design and implementation that goes beyond developing technologies and includes assessing the Manpower, Personnel, and Training requirements necessary to optimize performance. Mission effectiveness, whether for an organization or for an individual, will be minimal if the system is not properly integrated. In organizations that depend on technological systems as the primary to complete a task, true experts can determine exactly where breaks in the system occur and how to resolve these issues with a technological approach. However, when processes depend more on people than technology, there may be an increased likelihood for errors in judgment. Additionally, there may be fewer opportunities for quality control and systematic indicators that a problem exists. The key to risk management prior to making a decision is being able to balance uncertainty with action.

NATURALISTIC DECISION MAKING AND MACROCOGNITION
Adapting to the changing environment and thriving therein may seem ideal for experts who are successful, but the reality is most experienced experts learn the intricacies of their craft during crisis, or even failure. In many instances, some falter in chaos to the point of mission ineffectiveness. How we overcome these delays is a result of macrocognitive concepts, which can be summarized into two groups: functions and processes. The four functions are decision making, sensemaking, insight, and complex learning. The four processes are detecting problems, managing risk, managing uncertainty, and coordinating.

We recommend further investigation into risk management to identify how various groups asses work processes. Organizationally, risk guides the way groups code activities, decide on real-time and continuous processes, and assess outcomes within the construct of any activity – understanding risk is vital to organizational success. When accomplishing tasks in highly stressful, no-fail environments, teams may depend on the use of technology to supplement and verify tasks. However, the human remains a part of the loop regardless of the level of technological depth. Accordingly, we hold the power of human decisions as the focus of this analysis.

DISSECTING RISK MANAGEMENT
Through personnel training, assessment and compensation, organizations develop the schema for how they value individuals based on their ability to accomplish tasks. The social and experiential aspects of the workplace make it difficult to codify how a person compares his or her performance to another. Therefore, a person's ability to assess those around them will feed into his or her assessment of risk; ultimately this risk analysis guides how much or how little we employ technologies to overcome human deficiencies. In simpler terms, when one lacks confidence in the capabilities of another, he or she may prefer to use a technological approach

and bypass that person altogether. This can be problematic when resources are limited and time is an issue. Further, if the technology fails, significant energy could be wasted in finding a work-around solution. It is necessary to further analyze risk as an interaction between four constructs : fear, failure, planning, and productivity. The following sections detail each of these as a recommended subset of risk management within macrocognition. Further, one of the Army's most complex current issues, how to assist Soldiers who are in extreme distress, serves as a foundational topic for this construct.

Fear of Unknown Knowledge

In its simplest form, fear is synonymous with the stress that results from a lack of information. In a knowledge vacuum, people tend to assume the complete worst or the absolute best, instead of the most likely outcome. This void disconnects the individual and the system, because he or she replaces logic with the stress response and behavioral outcomes that counter progress in accomplishing the necessary tasks. At times, our assessment of available knowledge may be based upon our lack of confidence in the organization's knowledge management processes or our assessment of other team members' capabilities.

Assessing Personal and Collective Failure

It may be hard for a person to separate another person's capabilities from his or her potential failures. The two should not be synonymous, however how an organization classifies the lessons learned from past failures may guide how they relate individual skills to collective potential. High levels of self-efficacy in a collective group can make even the most inexperienced teams' mission effective. However, a team member may be unaware of how another person has performed in the past and not knowing whether the person has been successful in the past may hinder the relationship between the two. Additionally, in a worst-case mission failure scenario, the only way a person measures success is by being fully aware of how their supervisor will react.

Thinking while Planning

The natural tendency to plan for the worst and best case scenarios is not only a skill, but an art developed through experience. Because our thinking normally follows the cognitive path developed from years of constructing a schema around what works and what does not, some seldom venture from structured thinking (such as an outline) into creative thinking (such as a concept map). Inability to think outside the box during the planning process plagues the worker who aims to be busy, but may not be effective. Those who value "executing" over "planning" will sacrifice the time it may take to make a calculated decision for the short-term gain of making a decision at all.

Productivity while Working

When accomplishing an organizational task analysis, we may find the assessment biased by what the assessor considers productive, especially if the job being assessed is not one he or she themselves perform. In other words, it is very easy to seem busy or stagnant when the person making the assessment is unaware of the job's steadystate. In assessing whether others are "busy enough," we may overlook the fact that many factors of the job and personnel may be grossly under- or over-stated. Some will delay in accomplishing a task because they are willing to wait on the entity that seems busy but really is not working (and therefore over-valued) in order to wait on the entity that is over-tasked but not equipped to handle the workload (and therefore under-valued). In instances where a complex problem requires the assistance of multiple parties, this can be extremely problematic.

MANAGING RISK IN A CHALLENGING LEADERSHIP ENVIRONMENT

The proposed construct is especially important in any process that relates to the Human Resources (HR) field where the primary source of information and work hours comes from people instead of technology. HR in the Army continues to evolve to increase its technological capabilities, but maintains human interaction as a fundamental requirement for HR operations. The HR system most often interacts with its "expert" in the fundamental act of an Army leader (the expert) taking care of his or her soldier (the customer) using any of the personnel services available. A 2013 study included interviews with 24 active-duty Army soldiers who provided feedback on the Army Suicide Prevention Program. The researchers concluded a majority of the Soldiers understood the value and emphasis leaders place on the program, but did not trust the training construct to be the best line of defense when responding to suicidality. Further, the program's training emphasis of identifying suicide risk factors and depending on the 'buddy system' for identification of these factors were not rated as important to the interviewees as leader engagement and increasing personal protective factors associated with help-seeking behaviors.

We venture to guess the backdrop of the entire risk management construct is organizational trust (or lack thereof), which is key to our discussions on mission command and human systems integration. These proposed components of risk management reiterate the need for leaders to know how to seek, understand, and employ a response to a sexual assault or suicidal ideation at a moment's notice. Without adaptation from the "norm" many leaders would not know how to respond in order to provide vital assistance to those soldiers and family members in need.

First, there must be an accurate assessment of the knowledge gap or the leader will resort to the so-called "fear tactics" approach to providing help. This approach abandons discovering what information is known and focuses on the information that is not. Specifically, soldiers may not feel inclined to disclose the reasons for suicidal ideations, but leaders cannot be so fearful of the reasons that they miss an opportunity to provide assistance. Good leaders know that trying to force a person or a provider into this type of aid can be isolating and counterproductive.

Next, there must be a focus on defining success and failure, or an individual failure may be misconstrued as an organizational one. Ideally, individual success will be championed by the organization, but not overshadowed as an organizational win in all instances. A leader who identifies that a soldier needs help will build a plan around the individual's view of success, both in the short and long-term. The role of the leader is to find solutions that are best for all involved, as opposed to employing solutions that only avoid their personal failure.

Third, there must be a time-efficient planning process that does not end with the individual wasting unnecessary time due to a stifled thought process. The planning process is continuously adjusted based on updates to resource estimates and the maturity of the problem. In instances of emotional distress, there may not be an available clinical solution, but there may be an opportunity to help someone come up with a plan that addresses his or her basic needs. The expert leader accepts a changing plan over a failing one. Finally, this relates directly to being productive at all times, working to find solutions despite our limited resources and unlimited number of tasks.

CONCLUSION

Although this construct is not new with respect to naturalistic decision-making, we propose a closer examination of how people make decisions based upon how they assess risk. Within soft systems, a person becomes the gatekeeper of information and communication, not a computer. The decisions people make within soft systems are complex and evolving, and most importantly, time-sensitive. As both designers and users of the systems, experts must be aware of their own understanding (metacognition) and constantly assess collective adaptability (macrocognition) or few will be capable of implementing necessary system changes.

We continue to develop the conversation of how to best train our soldiers. In today's austere environments and complex matters, we must provide training and assessments that discuss risk management from the macrocognitive perspective. Most individuals prefer human solutions to human problems, understanding that technology is important but still unable ont make decisions on its own. Soldiers trust the leader who approaches the unknown willing to take a calculated risk, but avoid leaders who respond to complexity unwilling to accurately assess associated risks. As we continue to discuss the Human Dimension as a combat multiplier in military operations, further understanding of risk management and decision making is imperative.

REFERENCES

Bell, B. (2013). *A Human Systems Integration analysis of the Army suicide prevention program* (Master's thesis). Monterey, California: Naval Postgraduate School.

MacroCognition LLC. (2014). *Overview and macroCognitive framework,* Retrived from http://www.macrocognition.com

Assessing Hitting Skill in Baseball using Simulated and Representative Tasks in the Laboratory

Patrick K. BELLING[a,b], Jason SADA[b] and Paul WARD[c]
[a]*Michigan Technological University, Houghton, MI, USA*
[b]*Axon Sports, Scottsdale, AZ, USA*
[c]*University of Huddersfield, Huddersfield, West Yorkshire, UK*

ABSTRACT
Previous research has demonstrated that the ability to accurately anticipate the outcome of dynamic and representative situations in laboratory settings is an effective predictor of skill-level in many sports (for a review, see Ward, Williams, & Hancock, 2006). Other researchers have demonstrated that speed, in addition to accuracy, is an important component of skilled performers in sport (e.g., Jones & Miles, 1978; Savelsbergh, Williams, Van Der Kamp, & Ward, 2002). The current research aims to leverage this body of research in developing and evaluating a commercially available software tool designed for the assessment of such sports skills developed by Axon Sports. In this research we use the Axon tool to assess situational anticipation skill in an NCAA Division 1 baseball team. The results provide support that anticipation accuracy and speed are useful indicators of skill in sport and extend the application of this body of work into a real-world setting.

KEYWORDS
Decision making; anticipation; sport

INTRODUCTION
In most complex and dynamic domains, especially sports, the ability to anticipate the actions of others is a necessity for making quick and accurate decisions, and for executing those decisions effectively. In football, for instance, a successful quarterback must proactively anticipate the type of play, such as a blitz or a particular coverage that their opponents will employ next in order to avoid using an overly reactive strategy. Likewise, a successful soccer goalkeeper must anticipate the direction of a shot prior to the foot of the striker kicking the ball, and a successful baseball hitter must anticipate the trajectory and speed of a pitch prior to the ball leaving the pitcher's hand, or risk not being able to reach, or hit, the ball in time before it crosses the goal line, or plate, despite executing a good decision. Frequently, such anticipations have to occur prior to any obvious start of play (e.g., the ball being snapped) or prior to more easily recognizable cues (e.g., ball flight in soccer and baseball) in order to maximize the chances of success within the available time window. While readily apparent in these sports examples, early and accurate anticipation is critical to successful performance in many dynamic and complex domains, including driving, aviation and surgery to name but a few (for a review, see Suss & Ward, 2015).

Several researchers have investigated athletes' skill at anticipating future actions of opponents using representative or simulated laboratory tasks, often by using temporal occlusion-based methods (for reviews, see Ward, Suss, & Basevitch, 2009; Suss & Ward, 2015). This method, similar to the SAGAT (Endsely, 1995) albeit with a much longer history (see Haskins, 1965; Ward et al., 2008), is used to present, near-first-person, video-based scenarios (e.g., unfolding patterns of sport play) to participants up until a particular point in the play (e.g., foot-to-ball contact in soccer, racket-to-ball contact in tennis) where the participant has to make a critical prediction or decision. At this critical moment, the stimulus is typically occluded from participant's vision (e.g., Ward, Ericsson, & Williams, 2013; Belling, Suss, & Ward, 2014) or the last frame of action is frozen on screen (e.g., Johnson & Raab, 2003)—without being given access to the actual outcome of the play—and the participant is asked to complete their task (i.e., predict the next action/move by their opponent; decide on a course of action for themselves; execute their preferred course of action, etc.). Others have adapted this method in the field using, for instance, liquid crystal occlusion glasses which are set to occlude vision during real-life tasks via a specific timing device triggered by a specific event, such as an aspect of ball flight (e.g., Starkes, Edwards, Dissanayakee, & Dunn, 1995) or by the actions of the participant (e.g., Oujedans & Coolen, 2003).

Across several studies, researchers have demonstrated that expert athletes are more accurate and/or faster than novices when anticipating the outcome of particular plays from their domain of expertise (i.e., specialist sport) (e.g., Abernethy, 1990; Abernethy & Russell, 1987; Burroughs, 1984; Williams & Davids, 1995). For example, Abernethy and Russell (1987) presented videos of badminton players hitting the shuttlecock from the viewpoint of an opposing player. The video footage was occluded at varying times around the moment when the opponent's racket hit the shuttlecock. Expert badminton players were able to anticipate the flight path of the shuttlecock more accurately than novice players. Subsequent analyses of eye gaze data revealed that expert players used more information from early in the action sequence than novice players. While novice players fixated on the racket of the opposing badminton player, experts fixated on the arm of the opponent in addition to the racket. Similar findings were presented by Abernethy (1990) when experts and novices anticipated squash shots.

In another study using the temporal occlusion method in tennis, expert and novice tennis players viewed videos of serves and were asked to identify where, on the court, the serve would land (Jones & Miles, 1978). Occlusion of these videos occurred at pre-, near-, and post-contact of the the server's racket and the tennis ball. Experts anticipated the location of the serve more accurately than novices, but this effect was more profound in the earlier occlusion conditions (e.g., pre- and near-contact) that forced participants to use rely on more subtle information in the action sequence that occurred prior to the point of racket-ball contact.

Further support of the early and accurate anticipation advantage of experts was demonstrated by Rosalie and Müller (2013). Karate athletes were categorized into expert, near-expert, and novice groups. Using specialized occlusion glasses, the athletes' vision was occluded during combat while anticipating and blocking the attacks of opponents. Occlusion occurred either after the attacking motion of the opponent began, after the initial head motion began, prior to any motion of the opponent and was compared to a condition in which no occlusion occurred. At each of these occlusion points, expert karate athletes were able to block attacks at a rate significantly above chance performance. Near-experts were only able to accomplish this when there was no occlusion or occlusion occurred after the attacking motion of their opponent. Novices were able to block attacks only when there was no occlusion.

Similar results were found among soccer goalkeepers. Savelsbergh et al. (2002) employed temporal occlusion methods to investigate the ability of soccer goalkeepers to anticipate the location of penalty kicks. Using a joystick, rather than whole-body physical response, expert goalkeepers were more accurate than novice goalkeepers, but responded later in the action sequence. While this may seem contrary to the line of research described thus far, Savelsbergh and colleagues also noted that experts made fewer corrective movements. In other words, expert goalkeepers confirmed their early anticipations with information presented later in the action sequence whereas novice goalkeepers reacted based on erroneous early information and, in general, had to correct more often based on later information long after the experts had responded.

In sum, across a number of studies, skilled athletes have been shown to anticipate the outcome of dynamic situations in their sport with greater accuracy and speed. Such findings offer a potential explanation as to why expert athletes are able to perform at a reliably superior level compared to their novice counterparts in related contexts in their natural ecology (for a review of transfer effects see Ward et al., 2006). However, it is likely that other perceptual-cognitive skills, such as recognition skill, may precede successful anticipation. This assertion is consistent with current descriptive and theoretical claims about intuitive decision making (see Klein, 1993). In the sport of baseball, in addition to investigating anticipation skill (e.g., capability to anticipate the end-location of the pitch, specifically the height and distance from one's body as it crosses the plate), a handful of researchers have also investigated the ability to recognize the type of pitch prior to release or in the early stages of the pitch trajectory (e.g., fastball, curveball, changeup, slider). Both are important skills for baseball hitters. In the context of training, Burroughs (1984) examined both pitch location and pitch recognition. Using a pretest-training-posttest design, Burroughs observed that athletes that received video simulations designed to train the ability to recognize and locate pitches performed better at these tasks (although not significantly so) than a control group that received no training. The training effect remained present in a six week follow up test.

More recently, Fadde (2006) investigated the transfer of training of these perceptual-cognitive baseball skills to hitting performance in a real game. NCAA Division I collegiate baseball players were placed into a training and control group that were ranked equally by the team's coaches. The training group engaged in video-based simulation training designed to improve pitch recognition and pitch location. Training was completed during a two-week period. Following the two-week training period, the team completed its 18-game pre-conference schedule games. During those games, the training group recorded a significantly higher batting average than the control group. The batting average is the number of hits for a given batter divided by that batter's number of times at-bat (i.e., number of times facing a pitcher in-game) and is a widely accepted metric of hitting skill in baseball.

Within baseball, a growing body of evidence has been accumulated which shows that such video simulation tools can be effective for training the requisite perceptual-cognitive skills for successful performance in the real-world (e.g., Fadde, 2006; Burroughs, 1984). It would seem logical that the assessment of these skills may be a powerful predictor of skilled performance as well, potentially offering a diagnostic tool capable of predicting skill deficiencies. In the current research we evaluate the relationship between performance on a video-based assessment of pitch recognition and location, as well as a zone hitting drill that required the use of both skills in tandem, and the skill-level of near-expert baseball players. To assess these perceptual-cognitive skills under standardized conditions, we use a technologically advanced and innovative new software package developed by Axon Sports. The assessment software not only presents participants with temporally occluded baseball pitches similar to previous research, but also automatically records accuracy and time of response using a specific mode of interaction. Given that our expectation is that these indices will provide a valid assessment of the requisite cognitive skills players' for superior batting skill, we hypothesize that accuracy will be positively related, and response time will be negatively related, to ratings of each players' hitting skill provided by the team's coaching staff.

METHODS

Participants. The participants in this research were 23 NCAA Division 1 baseball players. The players completed the Axon Sports Baseball Hitting Assessment (see below) from their native hitting stance (right-handed/left-handed). Eight batters completed the left-handed batter version of the assessment and fifteen completed the right-handed batter version. The assessment took approximately 20 minutes per participant. After completion of the assessment, participants received individualized feedback detailing their strengths and weaknesses.

Materials. The Axon Sports Baseball Hitting Assessment is composed of 162 video simulations. These video simulations were created using video footage filmed from the right batter box. Mirror image videos were created to display pitches as if the footage were filmed from the left batter box. This image flipping also flipped the handedness of the pitcher on-screen (e.g., a natural right-handed pitcher would appear to be a natural left-handed pitcher). The first pitcher, a natural right-handed pitcher (RHP), threw a combination of fastballs, curveballs, and changeups. The second pitcher, a natural left-handed pitcher (LHP), threw a combination of fastballs, sliders, and changeups. The third pitcher, a natural RHP, threw a combination of fastballs, curveballs, and changeups. The baseball assessments were completed on a 65-inch touch screen monitor.

Using the Axon Sports Baseball Hitting Assessment software, three separate hitting tasks were created. Pitch Recognition (PR) required participants to select the correct type of pitch (e.g., fastball) from among the three pitches thrown by a particular pitcher (e.g., fastball, curveball, changeup) by touching the area of the screen corresponding to that type of pitch in a multiple choice format. Pitch Location (PL) required participants to select from nine subzones, representing the strike zone in baseball, as to which subzone the ball would pass through when crossing the plate. Zone Hitting (ZH) presented participants with an area of the strike zone (four of the nine subzones) and a type of pitch (e.g., fastball). When the participants recognized that type of pitch heading into the highlighted area of the zone (i.e., both pitch-type and pitch-location criteria were met), they were instructed to press a button on the screen to indicate swinging at that pitch. All three tasks contained a high- and low-occlusion condition. Moment of release (MOR) is defined by the frame at which the ball leaves the pitcher's hand and is often used as a critical moment in this type of research (see Fadde, 2006). During the PR task, the high occlusion pitches were occluded at MOR and low occlusion pitches were occluded at MOR + 10 frames (i.e., 10 video frames after the designated MOR frame). During the PL and ZH tasks, the high occlusion pitches were occluded at MOR + 2 frames and the low occlusion pitches were occluded at MOR + 10 frames. This was done so a very slight indicator of the ball's flight path could be seen when locating the pitch was a requisite of the task.

Procedure. Before completing each of the assessment tasks, three calibration videos were shown on screen to facilitate the batter adjusting their stance and location facing the screen to maximize lifelikeness. Next, participants completed 54 video simulations for that pitcher. This included 18 simulations of PR, PL, and ZH each, respectively. Of these 18 simulations, 9 were completed at high- and low-occlusion, respectively. Low occlusion pitches always followed high-occlusion pitches because low-occlusion pitches contained more information than high-occlusion pitches. Seeing the pitches at low-occlusion would potentially aid the participant when viewing them at high-occlusion, whereas the opposite is much less likely. This procedure was completed for all three pitchers. Right-handed and left-handed batters saw identical but mirror-image pitches. Therefore, right-handed batters completed the assessment following a RHP-LHP-RHP format while left-handed batters completed the assessment following a LHP-RHP-LHP format. Once the assessment was completed, participants received highly detailed and individualized feedback on each of the tasks and pitchers to help identify their strengths and weaknesses on the assessment.

Analysis. The variables of interest are accuracy and response time averaged across all of the tasks, and skill ratings. Accuracy was defined as the proportion of correct responses. Possible accuracy scores ranged from zero to one. Response time was defined as the average response time per trial (i.e., the time from occlusion until an answer was selected) and was measured in milliseconds. Ratings of batting skill were provided by the coaches of the collegiate team who have extensive experience working with the players. Skill ratings ranged from five to one. Five indicated an excellent batter. Four indicated a good batter. Three indicated an average batter. Two indicated a below average batter. One indicated a considerably below-average batter. Because skill level ratings were not a continuous variable, Spearman's rank-order correlation coefficient was used to compare accuracy, time, and skill.

RESULTS

Descriptive statistics can be found in Table 1. Recall that we hypothesized that accuracy on these tasks and batting skill rating would be positively related. Additionally, we hypothesized that response time on these tasks and batting skill rating would be negatively related. In accordance with our hypothesis, accuracy and batting skill rating were significantly positively related ($\rho = 0.67$, $p < 0.01$). In contrast to our hypothesis, response time and skill were not related ($\rho = 0.18$, $p = 0.42$). Time and accuracy were also not related ($\rho = 0.03$, $p = 0.91$). When time was included as a covariate in the analysis of accuracy and skill, it was not a significant factor ($F = 0.05$, $p = 0.84$).

	Skill Rating	Accuracy	Response Time
Player 1	5	0.766	1.476
Player 2	5	0.754	1.587
Player 3	3	0.690	1.597
Player 4	5	0.673	1.696
Player 5	5	0.643	1.334
Player 6	2	0.649	1.838
Player 7	5	0.655	1.291
Player 8	2	0.661	1.317
Player 9	1	0.637	1.284
Player 10	3	0.661	1.794
Player 11	3	0.643	1.481
Player 12	2	0.626	1.684
Player 13	3	0.626	1.615
Player 14	2	0.608	1.708
Player 15	1	0.608	1.659
Player 16	4	0.614	1.807
Player 17	3	0.591	1.858
Player 18	1	0.579	1.389
Player 19	1	0.591	1.601
Player 20	1	0.579	1.167
Player 21	4	0.585	1.722
Player 22	1	0.550	1.573
Player 23	1	0.556	1.48
Mean	2.739	0.632	1.563
SD	1.544	0.055	0.196

Table 1. Batting skill ratings, accuracy, and response time.

Additional exploratory analysis revealed that response time and batting skill rating approached significance when analyzing only the data from the higher-ranked players (i.e., skill ratings of 3, 4, and 5) in the hypothesized direction ($\rho = -0.53$, $p = 0.08$).

DISCUSSION

Our hypotheses were partially supported. Accuracy on the Axon Sports Baseball Hitting Assessment was significantly and positively related to hitting skill in the real world, as rated by coaches who possess both expertise in the sport and familiarity with the players. This suggests that the ability to accurately recognize the type of pitch and locate the pitch in advance of crossing the plate—two perceptual-cognitive skills that can be assessed using representative simulation tasks under controlled laboratory conditions—are useful predictors of on-the-field skill, in addition to being contender skills for training designed to accelerate expertise (see Burroughs, 1984; Fadde, 2006). Assessments of these perceptual-cognitive skills may be useful for collegiate baseball teams seeking the top talent in hitting. Further work is needed to validate these types of tests at other skill levels (e.g., professional, semi-pro).

Counter to our hypotheses, response time was not significantly related to the coaches' ratings of batting skill. This could be because time to respond on the touch screen interface is a qualitatively different mode of responding compared to swinging a bat in the real world. However, it is important to note that among the higher rated (i.e., more skilled) players, this relationship approached significance. This suggests that speed may play an important role in higher rated players, whereas accuracy explains most of the variance among the lower skill levels. Further work is certainly needed to substantiate this claim, however. Future research should seek to establish the predictive power of accuracy and response time during simulated hitting tasks on skill among more elite players (e.g., professional-level baseball players). Future research should also consider the creation of a more real-world response measure that integrates speed and accuracy. In the natural ecology, athletes playing dynamic sports must anticipate situational outcomes accurately and quickly in order to obtain success.

In general, this research offers further support for the use of perceptual-cognitive skills as a predictor of real-world skill in sport domains (see Abernethy, 1990; Abernethy & Russell, 1987; Burroughs, 1984; Williams & Davids, 1995). Not only has this research provided further support for this concept, but it has also validated a temporal occlusion tool that is readily available for the sports industry. This research offers a rather straightforward design for bridging the gap from academia to more applied settings, particularly within sport. However, future research should seek to validate a similar approach not only in other sports, but also in other complex and dynamic domains where quick and accurate anticipation and decision making are critical to successful performance (e.g., military, law enforcement, driving).

ACKNOWLEDGEMENTS

We would like to thank Axon Sports for providing state of the art simulation tools for collecting the data presented in this paper.

REFERENCES

Abernethy, B. (1990). Anticipation in squash: Differences in advance cue utilization between expert and novice players. *Journal of Sport Sciences, 8*(1), 17-34.

Abernethy, B., & Russell, D. G. (1987). The relationship between expertise and visual search strategy in a racquet sport. *Human Movement Science, 6*(4), 283-319.

Belling, P. K., Suss, J., & Ward, P. (2015). Advancing theory and application of cognitive research in sport: Using representative tasks to explain and predict skilled anticipation, decision-making, and option generation behavior.*Psychology of Sport and Exercise*, *16*, 45-59.

Burroughs, W. A. (1984). Visual simulation training of baseball batters. *International Journal of Sport Psychology, 15*, 117-126.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(1), 65-84.

Fadde, P. J. (2006). Interactive video training of perceptual decision making in the sport of baseball. *Technology, Instruction, Cognition and Learning, 4*(3), 265-285.

Haskins, M. J. (1965). Development of a response-recognition training film in tennis. *Perceptual and Motor Skills*, 21, 207-211.

Johnson, J. G., & Raab, M. (2003). Take The First: Option-generation and Resulting Choices. *Organizational Behavior and Human Decision Processes 91*, 215-29.

Jones, C. M., & Miles, T. R. (1978). Use of advance cues in predicting the flight of a lawn tennis ball. *Journal of Human Movement Studies, 4*(4), 231-235.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G.A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp.138–147). Norwood, NJ: Ablex.

Oudejans, R., & Coolen, B. (2003). Human kinematics and event control: on-line movement registration as a means for experimental manipulation.*Journal of Sports Sciences, 21*(7), 567-576.

Rosalie, S. M., & Müller, S. (2013). Timing of in-situ visual information pick-up that differentiates expert and near-expert anticipation in a complex motor skill. *The Quarterly Journal of Experimental Psychology*, DOI: 10. 1080/ 17470218. 2013. 770044.

Savelsbergh, G. J. P., Williams, A. M., Van Der Kamp, J., & Ward, P. (2002). Visual search, anticipation, and expertise in soccer goalkeepers. *Journal of Sports Sciences, 20*, 279-287.

Starkes, J. L., Edwards, P., Dissanayake, P., & Dunn, T. (1995). A new technology and field test of advance cue usage in volleyball. *Research quarterly for exercise and sport*, *66*(2), 162.

Suss, J., & Ward, P. (2015). Predicting the future in perceptual-motor domains: Perceptual anticipation, option generation and expertise. In R. R. Hoffman, P. A. Hancock, M. Scerbo, and J. L. Szalma (Eds.), Cambridge handbook of applied perception research (pp. 951-976). New York, NY: Cambridge University Press.
Ward, P., Ericsson, K.A., & Williams, A.M. (2013). Complex perceptual-cognitive expertise in a simulated task environment. *Journal of Cognitive Engineering and Decision Making, 7*, 231-254.

Ward, P., Suss, J., & Basevitch, I. (2009). Expertise and expert performance-based training (ExPerT) in complex domains. *Technology, Instruction, Cognition and Learning*, *7*(2), 121-145.

Ward, P. Williams, A. M., & Hancock, P. A. (2006) *Simulation for performance and training.* In: The Cambridge Handbook of Expertise and Expert Performance. Cambridge University Press, pp. 243-262. ISBN 9780521600811

Williams, A. M., & Davids, K. (1995). Declarative knowledge in sport: A byproduct of experience or a characteristic of expertise? *Journal of Sport & Exercise Psychology, 17*(3), 259-275.

Revealing and Assessing Cognitive Processes Underlying Cultural Acuity Through Domain-Inspired Exercises

Matthieu BRANLAT[a], Julio C. MATEO[a], Michael J. McCLOSKEY[a] and LisaRe BROOKS BABIN[b]

[a]361 Interactive, LLC., Springboro, OH [b]U.S. Army Research Institute for the Behavioral and Social Sciences, Fort Leavenworth, KS

ABSTRACT

A Soldier's ability to develop an understanding of the sociocultural aspects of unfamiliar environments is critical to achieving mission success. In this research, performance-based methods were developed to assess a Soldier's ability to learn about, interpret, and adapt to unfamiliar cultural environments. Six complementary methods were designed with the goal of recreating key demands of unfamiliar environments and eliciting cognitive processes and behaviors similar to those required in foreign operational settings. A sample of U.S. Army Soldiers participated in this research. Data were analyzed to evaluate the potential utility of each method and inform revisions. Overall, the methods developed successfully elicited and captured relevant, observable behavior to assess cultural acuity. A framework was developed to better understand inter-method differences and complementary features. The findings serve as a foundation for the development of future performance-based batteries to assess cross-cultural competence and similar competences.

KEYWORDS

Decision making; military; cross-cultural competence; cultural acuity; sensemaking; assessment.

INTRODUCTION

During non-kinetic operations (e.g., military transition teams), U.S. military personnel often work with foreign civilian and military personnel to achieve a common goal. In these missions, Army leaders who have the ability to quickly and effectively develop a working understanding of the important sociocultural aspects of an unfamiliar environment are better equipped to succeed. This understanding enhances their ability to develop culturally sensitive courses of action to achieve mission success while simultaneously minimizing potential negative, unintended consequences of those actions. Over the past decade, the Department of Defense (DoD) has undertaken and sponsored numerous research efforts in an attempt to better understand (e.g., Abbe, Gulick, & Herman, 2007), train (McCloskey, Behymer, & Mateo, 2012), and assess (Gabrenya, Griffith, Moukarzel, Pomerance, & Reid, 2012) *cross-cultural competence* (*3C*) in operational settings and enhance the effectiveness of U.S. military personnel when interacting with individuals from diverse cultural backgrounds. DoD-sponsored 3C research has highlighted the importance of understanding, training, and assessing *general* 3C *knowledge, skills, abilities, and attitudes* (*KSAAs*) that apply across cultures (Abbe et al., 2007) and the need to move beyond *culture-specific*, 'smart-card' approaches focused exclusively on upcoming deployments. This culture-general approach is not only empirically supported, but also makes sense from a strategic standpoint since the exact location of the next conflict requiring U.S. ground troops cannot be predicted, but independently Soldiers will need to quickly and effectively make sense of unfamiliar cultural situations and adapt their behaviors. This research specifically targeted the development of performance-based methods to assess a Soldier's ability to interpret, learn about, and adapt to unfamiliar cultural situations to achieve mission success.

Understanding Cross-Cultural Competence and Cultural Acuity

The research presented here builds on a model of general 3C empirically derived from data collected from Warfighters, and which captures the field requirements of deployed military personnel (McCloskey, Behymer, Papautsky, Ross, & Abbe, 2010). The five-factor model is described in Table 1.

Table 1. The five factors in McCloskey et al.'s (2010) model of general 3C.

Factor	Definition
Cultural Interest	Willingness to learn about local culture and engage with local nationals as a way to accomplish the mission
Cultural Relativism	Awareness of cultural differences when dealing with individuals from diverse

	backgrounds, and openmindedness regarding unusual practices in other cultures
Cultural Acuity	Ability to develop effective working understandings in cross-cultural situations, even when the target culture is highly unfamiliar
Relationship Orientation	Tendency to value and show interest in personal relationships
Interpersonal Skills	Ability to present themselves in a way that promotes positive short- and long-term interactions

In this effort, the focus was on assessing cultural acuity, which is considered a key aspect of 3C in operational settings. To clarify the nature of cultural acuity and guide the development of assessment methods, the research team fleshed out the KSAAs comprising cultural acuity, emphasizing those KSAAs considered as most relevant to learning about unfamiliar cultural environments *through direct observation*. Table 2 describes the KSAAs identified as supporting the capacity to effectively observe one's environment, interpret environmental cues, and develop a functional understanding of the situation and effective courses of action to achieve mission objectives.

Table 2. KSAAs identified as critical to cultural acuity.

KSAA	Definition
Observation	The cognitive processes underlying an individual's ability to detect cues that provide useful information about the target culture (e.g., beliefs, values) when observing cross-cultural interactions
Perspective Taking	The cognitive processes underlying an individual's ability to step outside one's own cognitive viewpoint to understand how other people perceive, think, and/or feel in specific situations
Sensemaking	The cognitive processes (e.g., hypothesis generation and revision, information seeking) underlying an individual's ability to develop sensible explanations when faced with surprising or ambiguous stimuli
Cultural Awareness	An individual's capacity to recognize one's cultural biases and how they impact one's perceptions and assessments
Interpersonal Decoding	An individual's capacity to use another person's observable behavior to learn about their disposition
Cognitive Complexity	An individual's capacity and willingness to acknowledge that an issue can have many competing perspectives, to realize the links among them, and to conceptually integrate across them
Cognitive Flexibility	An individual's tendency to use broad, inclusive cognitive categories when thinking about the world and the ability to switch among these different categories

General Approach to Assessment and Design

This research effort used an unconventional approach to the design of methods to assess cultural acuity. Rather than isolating individual KSAAs and using responses entered by the participant to assess their level on each, the team developed a set of methods that targeted multiple KSAAs simultaneously from different perspectives and relied on observers rating participant performance to assess the participant's level of cultural acuity. There were two important influences that shaped this approach to assessment and design: *Naturalistic Decision Making* (Zsambok & Klein, 1997) and *Cognitive Systems Engineering* (Woods & Hollnagel, 2006). Ecological and cognitive validity constituted the main drivers behind the selection and development of assessment methods. That is, the primary emphasis during the development of the performance-based methods was (a) to reflect the demands of real-life situations that Soldiers face in operational situations in which cultural acuity is required and, as a result, (b) to elicit cognitive processes similar to those in which Soldiers engage in those situations. The starting point of the approach is therefore an understanding of the characteristics and demands of the operational world, guided by a conceptual framework of cognitive abilities in such environment. A set of performance-based methods was developed to assess cultural acuity as a whole from different complementary *perspectives* (cf. *Results and Discussion* section). Such an approach contrasts with typical assessment projects which tend to: (1) rely heavily on self-report and declarative knowledge, and (2) break down the object of assessment (e.g., cultural acuity) into components investigated in isolation. Another central assumption of the project was the need to assess the quality of the cognitive processes in which participants engage while completing the methods (*process*), rather than the accuracy of their responses (*outcome*). Such focus stems from the team's understanding of the fundamentally dynamic and cyclical nature of cognitive processes such as sensemaking: situations tend to unfold over time, evidence becomes available progressively, and new evidence sometimes conflicts with prior understanding. The methods proposed in this report were specifically designed to reveal and assess how people build an understanding of culturally challenging situations over time through seizing opportunities to gather more evidence and making sense of it. Assessing the quality of the process underlying cultural acuity from observable behavior nonetheless presents important challenges (e.g., scoring cannot consists of comparing responses to a known answer, but relies on more subjective assessments). A cycle of developments and revisions was followed over the course of the effort. Based on previous experience and knowledge, an initial set of methods (i.e., a first prototype) was developed. Initial feedback was gathered from in-house colleagues who were not familiar with the project and the methods were revised based on the resulting data. The data collection described below provided an opportunity to use the revised prototype to gather data from U.S. Army personnel. This data collection, in turn, provided substantial insight into several major aspects of the methods: the relevance and scope of the material, their usefulness to assessing processes underlying cultural acuity, challenges and opportunities for the administration of the various methods, and requirements for scoring. Findings from the data collection were then used to revise the methods further and produce a more focused, balanced, and administrable assessment battery.

METHOD

The following subsections describe the performance-based methods that were developed, summarize the design and findings of the data collection, and discuss the implications of the findings both for revisions of the assessment battery but also for the assessment of 3C and other similar competences using performance-based methods.

Candidate Assessment Methods

To guide the development of the assessment methods, the team identified a set of criteria that each of the resulting assessment methods would have to meet to be successful given the envisioned application setting:

- It elicits relevant observable behaviors that vary across participants as a function of their cultural acuity.
- It can be administered by a single administrator during a one-on-one meeting.
- It is self-contained (i.e., instructions include all guidance or training needed to administer the method).
- Administrator does not need extensive training or prior experience (i.e., any unit member could run it).
- It can be scored in real-time, without the need to record the sessions or analyze them after the fact.
- The whole assessment battery can be administered within a 2- to 3-hr period.

The six methods developed are described in Table 3. The potential of these candidate methods to support cultural-acuity assessment was investigated in the data collection described in the next subsections.

Assessment Method Description *Fictional Culture* Participants watch a video showing a group of actors acting out a meeting in an unfamiliar (fictional) culture. *Exercise* The video is stopped at certain points and participants are asked questions regarding the events, individuals, and culture in the video.

Unfamiliar Sport Participants watch a video showing two teams playing a match of a real sport that is most likely unfamiliar to *Exercise* participants. They are asked to try to learn as much as they can about how the sport works (e.g., rules, scoring) and to think aloud as they watch and control the video.

Table 3. Candidate assessment methods developed for this research effort

Dynamic Location Participants are virtually placed in an undisclosed location and asked to determine where in	the world they were
Exercise placed. The	program displays scenes from locations around the world, shown from the participant's point of view. Participants can control the interface to move, look around, or zoom in on objects of interest. They are also asked to think aloud as they complete the task.
Static Scene Exercise	Participants examine a series of photos from operational environments. They are asked to point out elements in the scene that they consider relevant to culturally assess the region and explain how those elements would affect their assessment.
Simulation Interview	Participants are presented with a developing scenario. After each new event is introduced, participants are asked a series of questions about how they would interpret the situation or what they would do given the circumstances.

Past Experience Participants are asked to recall relevant incidents from their own life in which they experienced certain *Interview* situations (e.g., moving to a new area). Once they provide an incident, participants are asked questions about their expectations, thoughts, and actions in those situations.

Participants and Procedure

A total of 34 U.S. Army Soldiers were recruited through the U.S. Army Research Institute for the Behavioral and Social Sciences and participated in the data collection. The sample consisted of 29 men and 5 women, ranging from 20 to 48 years old (M = 28 years, SD = 7 years). Soldiers included both officers and enlisted Soldiers, ranging in grade from PFC to CPT. They had served in the U.S. Army for an average of 6 years (SD = 6 years), totaling an average of 18.2 months of deployment (SD = 17.2 months). All sessions were scheduled for 90 min and took place in a classroom setting. At the beginning of each session, the administrator greeted the participant, briefly explained the purpose of the research, and asked for his or her consent to participate. All performance data were kept anonymous and cannot be linked to individual Soldiers. All sessions were audio recorded in their entirety for further analysis.

Qualitative Analyses

Recordings were fully transcribed and the research team subjected the resulting transcriptions to thorough qualitative analysis. The specific procedures used to examine the data varied from method to method to accommodate for method idiosyncrasies. However, analyses for all methods examined:

- *Response variability:* whether responses showed variability across participants.
- *Response relevance:* whether individual differences appeared to reflect differences in cultural acuity.
- *Manifestation of cultural-acuity KSAAs:* whether KSAAs underlying cultural acuity identified earlier in the research process were manifested in the participant responses to different methods.
- *Method revisions:* Potential modifications that could result in increased response variability and relevance, or in reductions of overall administration time (e.g., redundant or unclear questions).
- *Scoring development:* Potential techniques to enable administrators to score methods in real-time.
- *Supports for inexperienced administrators:* Potential revisions to the administration and scoring guides to enable individuals with no previous experience (e.g., military unit member) to administer the methods.

Typically, analyses involved tasking members of the research team with reviewing and scoring data in terms of their estimated level of cultural acuity from 1 (low) to 5 (high). For each of the participants, raters also wrote their rationale for the score given. Raters then met to compare their ratings, discussed the rationale for their ratings, identified inconsistencies, and proposed a scoring guide to be used in a more systematic manner. Discussions also resulted in the development of a list of cues and strategies used by participants, which was eventually incorporated into the scoring guide to support inexperienced administrators. Qualitative analyses were also used to determine which KSAAs of cultural acuity were reflected in the think-aloud protocols.

RESULTS AND DISCUSSION

Overall Findings Across Methods

All six methods were received positively by participants and showed potential for supporting the assessment of cultural acuity in Army personnel. Analyses of responses supported the idea that KSAAs underlying cultural acuity were reflected in the data collected. Methods were revised based on the findings from qualitative analyses. Revisions included the elimination of questions whose responses showed low variability across participants, were unclear to participants, or were redundant with other questions. Other revisions involved more substantial changes to an individual method to address unanticipated challenges identified during the data collection. Qualitative analyses were also conducted to guide the development of scoring guidelines. Scoring guides presented administrators with a behaviorally anchored rating scale for each of the questions and/or trials within each method. The research team also developed note-taking supports to guide the attention of administrators during the scoring process. A framework with multiple feature dimensions was developed to classify and distinguish the properties of different methods. Individual methods were typically inadequate to assess *all* of the KSAAs underlying cultural acuity, but each of the methods was capable of supporting the assessment of at least a subset of KSAAs. Together, the six methods provided complementary perspectives that contributed to a comprehensive assessment of cultural acuity (cf. *Assessment Through a Battery of Performance-based Methods* subsection below). Next section illustrates through one of the exercises how the collected data was used to evaluate and revise the assessment battery. Although specifics vary, the description is representative of the design process for all methods, as well as of the general nature of exercises and evaluation.

A Closer Look at the Dynamic Location Exercise

The Dynamic Location Exercise (see Table 3 for a short description) aims at eliciting behaviors that are informative regarding the observation skills (e.g., picking up relevant cues) and sensemaking processes (e.g., information seeking, hypothesis generation and revision) that the participant is likely to display when faced with unfamiliar environments. The reaction to the Dynamic Location Exercise was overwhelmingly positive among participants. Overall, they found the task interesting and challenging, showed engaged behavior, and were motivated to figure out the locations. In fact, some participants even asked whether they could "play with it some more at home" to get better at it. The design of the Dynamic Location Exercise allowed researchers to gain access to the processes underlying behavior in this task and to reveal differences in performance. For example, participants differed in the extent to which they (a) used prior knowledge impacting recognition of relevant cues (e.g., style of taxis in England), (b) used of exploration strategies (e.g., seeking for highly informational cues such as street signs), (c) were able to form coherent hypotheses based on the integration of cues gathered, and (d) were able to test and revise hypotheses in the face of contradictory information. Unexpected design issues were identified during the data collection. For example, an unanticipated consequence of giving participants full freedom to move in any direction was that, once each trial began, the exact stimuli experienced by participants during the same trial differed substantially depending on their navigation choices. Importantly, navigation choices during the first few moves within each trial were not always strategic in nature, but rather the result of arbitrary exploration (not information seeking *per se*). Another unanticipated issue was the presence of signs that unequivocally revealed the location. While the research team attempted to prevent participants from accessing this type of information, the ability of participants to move freely in any direction made it impossible to completely eliminate these 'give-away' signs. As a result, some participants developed deliberate strategies consisting primarily (or even exclusively) of looking for these types of 'give-away' signs to complete the exercise. While such a workaround was often effective at accomplishing the stated goal of the method (e.g., determining where in the world the location is), it seriously hindered the administrator's ability to assess how participants interpreted other (less informative) cues in the environment during the process and, therefore, it was considered suboptimal for assessment purposes.

Revisions

The Dynamic Location Exercise was revised to address some of the unanticipated issues mentioned above. There were two main modifications: a restriction of exploration capabilities and a re-design of locations and sublocations to instantiate specific challenges associated with cultural acuity. Regarding the restriction of exploration capabilities, the revised version did not allow participants to move freely from the starting point. Instead, each trial contained three carefully selected sublocations within which participants could only look around (rotate) and zoom into any region of interest, but not move down the street (translate). Furthermore, the three sublocations were made accessible (unlocked) in a progressive manner. Once all were unlocked, participants could move back and forth between sublocations to explore each further or compare across them. Regarding the re-design of locations and sublocations, significant effort was invested to identify potential locations and sublocations for the revised version so that all 'give-away' signs were eliminated and, as a whole, a diverse set of characteristics and associated challenges were encountered in the exercise. For instance, the locations varied in richness and specificity of cultural information: one location only included rather generic cultural cues (e.g., flat areas, corn fields, rural), whereas another one included many rich and complex cultural cues (e.g., mix of cultures, religions, ethnicities). Some trials were designed so that progressive sublocations provided additional, consistent data to support participants' early interpretations, while other trials instantiated *garden path* problems in which "an initial setup that suggests one hypothesis [was] followed by a dribbling of contrary cues that indicate a different hypothesis" (Klein, Moon, & Hoffman, 2006, p. 72). These revisions are expected to enable administrators to better observe and qualify participants' sensemaking processes.

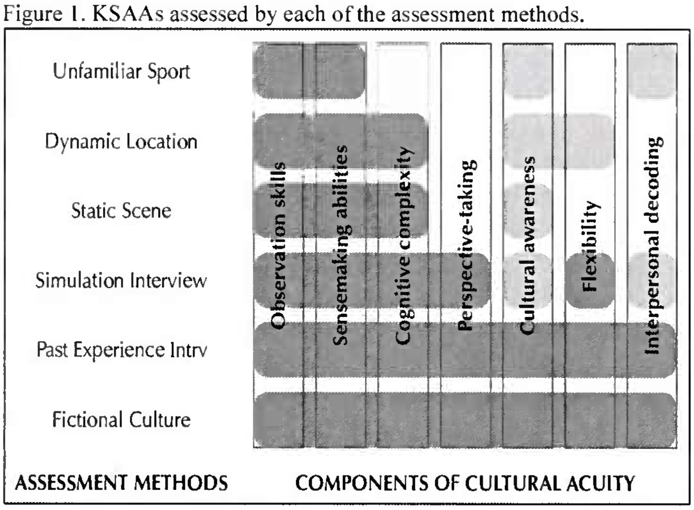
Scoring Development

The Dynamic Location Exercise appeared best suited to gather information about observation skills, sensemaking skills, and cognitive complexity. The capacity of participants to notice a variety of relevant cues in the locations (observation skills) could be rated according to both the quantity and the diversity of cues observed and scoring combined both aspects. In order to support the identification of cues considered by participants during verbalizations in real-time, data were used to devise classification of cue types, for instance relating to the natural environment (e.g., vegetation) or to people (e.g., language). The list was expected to be particularly useful for inexperienced administrators. The amount and diversity of available cues differed substantially across locations. While some locations afforded more observations of agricultural landscapes, others concentrated on built environments. For each location, the generic classification was therefore tailored to create a location-specific scoring guide highlighting those cue types that were especially relevant for making appropriate guesses for that location. The general category of sensemaking skills was divided into more tractable skills: hypothesis generation, information seeking, and hypothesis revision. Descriptions of differing levels of sensemaking skills in the context of the Dynamic Location Exercise were also developed to support administrators. For each location, the scoring guide included descriptions of anticipated cognitive challenges to sensemaking specific to that location in order to direct the attention of administrators to relevant aspects and facilitate scoring. Finally, cognitive complexity was manifested by the diversity of cues that a participant reported while completing the exercise as well as the demonstrated integration across cues. The consideration of cognitive complexity appeared particularly relevant to capture important aspects of responses in locations with richer and more diverse cues. The scoring guide for cognitive complexity consisted of a detailed description of expected types of behaviors and associated scores.

Assessment Through a Battery of Performance-based Methods

Candidate methods were designed to approach the same or similar phenomena from slightly different perspectives, using methods that differed in key features. The KSAAs underlying cultural acuity (see Table 2) were considered a useful framework to integrate findings across methods, describe the complementary nature of individual methods, and organize the scoring of the whole assessment battery. Even though methods could potentially reflect other KSAAs, the task of rating every answer in real-time on all seven KSAAs was considered too overwhelming for a single administrator. Instead, in most methods the top two or three KSAAs were chosen to be the focus of the assessment, based on how well suited the method was to assess those KSAAs. Figure 1 shows the KSAAs assessed using each method. The semi-transparent blocks represent KSAAs for which the method provides some information, but which were not targeted by design when used within the battery.

Figure 1. KSAAs assessed by each of the assessment methods.



In addition to the KSAAs addressed, other differences across methods are relevant to understand their complementary nature. We used six dimensions to characterize the methods and highlight the idiosyncrasies of each, their potential limitations, as well as the richness of the full assessment battery. These dimensions were: *level of interaction*, *nature of performance*, *perspective*, *dynamicity*, *domain/task fidelity*, and *relative richness*. The first three dimensions relate to the nature of the methods, whereas the last three are related to aspects that make the method more or less complicated. Figure 2 visually represents of how the various methods relate to each other and, as a whole, cover the space of possibilities across those dimensions.

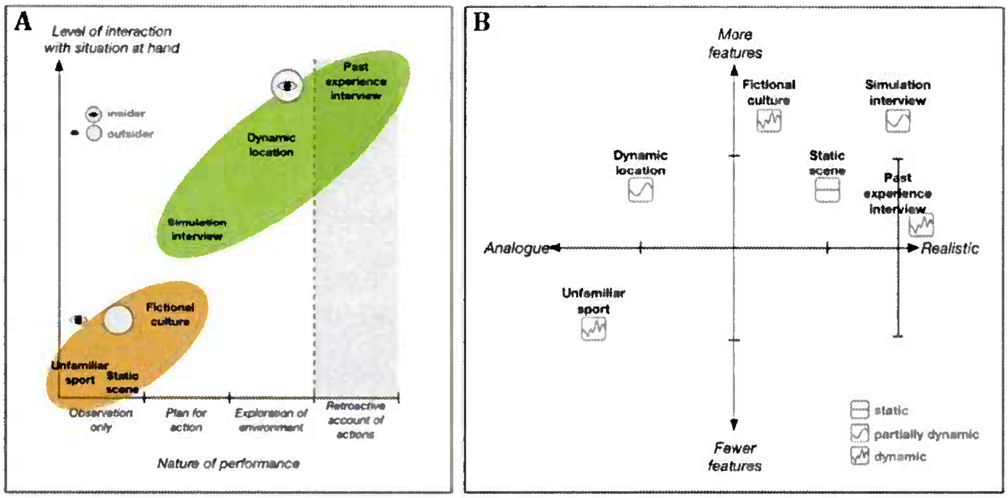


Figure 2. Assessment methods plotted relative to their characteristics. **A** (left): level of interaction, nature of performance, and perspective; **B** (right): relative richness, domain/task fidelity, and dynamicity.

The first noticeable feature of representation 2A is that methods appeared clearly divided between those involving an outsider perspective and those involving an insider perspective. It appears as if level of interaction was correlated with perspective: methods with insider perspective tend to have higher levels of interaction. While one would expect some level of correlation, there are also certain design decisions that could make methods with an outsider perspective more interactive. For example, the initial version of the Unfamiliar Sport Exercise included more interactive features (e.g., ability to pause, rewind) than the revised version. During the development of the battery, however, these features were eliminated to enable the assessment of more realistic information seeking behaviors during cross-cultural situations in which the participant would not have this level of control. Figures 2A and 2B both highlight the diversity of characteristics of the candidate assessment methods. They also reveal that some areas in the representation spaces remain uncovered, suggesting potential directions for future development of the assessment battery. For instance, Figure 2B suggests that an assessment method presenting high domain fidelity but fewer features might be useful when assessing novice participants.

Limitations and implications for future work

Establishing construct validity to the levels that are typically expected of assessment methods is challenging for these types of performance-based methods. The methods were designed to predict performance during deployments or other cross-cultural experiences, rather than to assess the underlying theoretical constructs *per se* (e.g., cognitive flexibility). That said, some level of convergent validity is expected when comparing the scores from this battery to scores from conventional assessment methods targeting individual KSAAs underlying cultural acuity. This could be one promising direction for future battery validation efforts. One limitation of the assessment battery, as currently designed, is that a participant's verbal ability is likely to influence the assessed level of cultural acuity obtained, since the resulting score relies heavily on the participant's verbal output during performance. While non-verbal performance measures are used to complement verbal measures in some methods, their contribution to the final score remains limited at this time. Future research should explore more non-verbal performance measures and investigate how performance in those measures may provide information about the processes that participants are following as they complete the tasks. In addition to difficulties in accurately scoring performance associated with specific exercises or aspects, one critical question is how to combine the various scores into a coherent global performance assessment. Given the variety of exercises and tasks within exercises, a simple averaging of scores across exercises is likely to hide significant variability in performance. Further efforts are required in order to provide more meaningful information to the administrator and more effective feedback to the Soldiers. The approach currently favored by the team involves constituting a rich performance profile based on the scores obtained in the various components of the assessment battery (e.g., see McCloskey et al., 2012). As mentioned earlier, the KSAAs can serve as a strong theoretical basis for the constitution of such scoring system.

As currently designed, scoring of participants requires the presence of an administrator who is simultaneously involved in the facilitation and real-time scoring of the individual exercises. Automating or supporting some of the administration or scoring methods through technology could reduce workload and help the administrator focus on the more important tasks (e.g., those who really require human judgment). Rather than automating scoring *per se*, technological tools could help keep track of specific observables and quickly fill scoring grids using those data. Some of the methods (e.g., Dynamic Location Exercise) are more conducive than others to such use of technology, given the higher amount of expectancies associated with performance at this method. The assessment battery was designed to serve as a tool to assess cultural acuity in a single session. However, other uses of such methods can be imagined. A first candidate would be to use the assessment battery to evaluate Soldiers' progress in the context of training and deployments. Comparing assessments over time would provide invaluable information about the extent to which the assessment methods developed here are effective at predicting who will perform better during deployments, as well as about the effectiveness of the cross-cultural training they receive. Because the methods were developed to capture and elicit the demands of real-world situations, the stimuli and tasks in those methods can provide a strong foundation to train skills that will be useful in operational settings. The provision of feedback to participants was not considered desirable in this assessment effort. However, future research should investigate how to develop and provide formative feedback to support the training of cultural acuity. The team is currently working on adapting a number of the exercises to support the development of a general 3C curriculum for special operators.

SUMMARY AND CONCLUSION

The research investigated the use of performance-based methods to elicit cognitive processes and observable behaviors similar to those encountered in operational situations as a way to assess cultural acuity in U.S. Army Soldiers. Six candidate assessment methods were developed, evaluated, and revised. The findings demonstrated the potential of performance-based methods designed to recreate the demands of operational situations to support assessment of Warfighters' cultural acuity. The findings also confirmed the feasibility and relevance of an approach based on a battery of complementary methods representing different overlapping perspectives, each only partially sufficient to assess cultural acuity. While work to develop appropriate formative feedback for these exercises is still underway, the exercises developed during this research effort also show promise as training materials to enhance cultural-acuity learning and performance. The research described in this paper represents a critical step and strong foundation in the development of performance-based methods to train and assess cultural acuity of Warfighters.

ACKNOWLEDGMENTS

This research was supported by the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI), Fort Leavenworth, KS, under contract W5J9CQ-13-C-0006. The views expressed in this article are those of the authors and do not necessarily represent the view of the Department of Defense. We thank all of the U.S. Soldiers who participated, as well as the ARI researchers who generously contributed to this effort.

REFERENCES

Abbe, A., Gulick, L. M. V., & Herman, J. L. (2007). *Cross-cultural competence in Army leaders: A conceptual and empirical foundation* (SR 2008-01). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. DTIC# ADA476072

Gabrenya, W. K., Griffith, R. L., Moukarzel, R. G., Pomerance, M. H., & Reid, P. (2012). Theoretical and practical advances in the assessment of cross-cultural competence. *Proceedings of the 2nd International Conference on Cross-Cultural Decision Making*, San Francisco, CA, 2911–2920.

Klein, G., Moon, B., & Hoffman, R. R. (2006). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems, 21,* 70–73.

McCloskey, M. J., Behymer, K. J., & Mateo, J. C. (2012). *CultureGear: Training cross-cultural perspective taking skills* (Final Technical Report). Arlington, VA: Office of Naval Research.

McCloskey, M. J., Behymer, K. J., Papautsky, E. L., Ross, K. G. & Abbe, A. (2010). *A developmental model of cross-cultural competence at the tactical level* (TR 1278). Alexandria, VA: U. S. Army Research Institute for the Behavioral and Social Sciences. DTIC# ADA534118

McCloskey, M. J., Behymer, K. J., Papautsky, E. L., & Grandjean, A. K. (2012). *Measuring learning and development in cross-cultural competence* (TR 1317). Fort Belvoir, VA: U. S. Army Research Institute for the Behavioral and Social Sciences.

Woods, D. D., & Hollnagel, E. (2006). *Joint cognitive systems: Patterns in cognitive systems engineering.* Boca Raton, FL: Taylor & Francis/CRC Press.

Zsambok, C. E., & Klein, G. (Eds.) (1997). *Naturalistic decision making.* Mahwah, NJ: Lawrence Erlbaum Associates.

HELP: Formalizing Frames in a Story of Sensemaking

Kevin Burns
The MITRE Corporation[7], *kburns@mitre.org*

ABSTRACT
"Frames" have been theorized in studies of sensemaking, but have not been formalized in a manner that can measure how well humans make sense of uncertain information. Here I use Bayesian concepts of hypotheses, evidence, likelihoods, priors, and posteriors (HELP) to define the components of frames and to model the dynamics of framing and reframing. This Bayesian approach is applied to a real-world story about one analyst's sensemaking, and used to identify several distinct types of reframing in the narrative account. The results were used as a basis for designing laboratory experiments to measure human performance in prototypical tasks of intelligence analysis, including cognitive biases relative to normative standards. Insights from these experiments, along with case studies obtained from practicing analysts, suggest the Bayesian approach used in this research can be applied as a structured analytic technique – to improve the rigor of naturalistic sensemaking in the field of intelligence analysis.

KEYWORDS
Sensemaking; mathematics and statistics; uncertainty management; judgment and decision making.

INTRODUCTION

Recent research on sensemaking has moved from conceptual theories (Klein, Moon & Hoffman, 2006a, 2006b; Klein, Phillips, Rall & Peluso, 2007) to computational models and empirical measures. In particular, the IARPA (Intelligence Advanced Research Projects Activity) program ICArUS (Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking) developed neural-computational models of human sensemaking (IARPA, 2010), and conducted laboratory experiments with human participants to test and evaluate the models (Burns, Bonaceto, Fine & Oertel, 2014). These experiments employed challenge problems (Burns, Greenwald & Fine, 2014; Burns, 2014b) designed to achieve a balance of empirical rigor and practical relevance.

For rigor in laboratory experiments, ICArUS (IARPA, 2010) required that human sensemaking be scored numerically as a percentage of theoretically optimal performance. This was to measure cognitive biases and to assess how well models could replicate human behaviour. For relevance to real-world intelligence, the experimental challenge problems involved prototypical tasks of geospatial analysis. These tasks were patterned after case studies of sensemaking obtained in interviews with practicing analysts.

Drawing on an existing *"data-frame theory of sensemaking"* (Klein et al., 2007), ICArUS challenge problems were designed to address the core processes of *"framing"* and *"reframing"* whereby *"data"* are explained in *"frames"*. This was accomplished by analysing a data-frame story of sensemaking (Klein et al., 2007), using Bayesian concepts to formalize the structure of frames and the nature of framing and reframing (Burns, 2014a) – as needed to measure human performance and cognitive biases in sensemaking experiments.

Here I outline these Bayesian concepts, apply them to the story of sensemaking, and explain how a Bayesian approach can be extended beyond ICArUS experiments to improve the practice of intelligence analysis.

METHOD

A Data-Frame Theory of Sensemaking

According to Klein et al. (2007), *"The data-frame theory postulates that elements are explained when they are fitted into a structure that links them to other elements. We use the term frame to denote an explanatory structure that defines entities by describing their relationship to other entities."* The associated processes include: *"The initial account people generate to explain events. The elaboration of that account. The questioning of that account in response to inconsistent data. Fixation on the initial account. Discovering*

 This publication is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA) ICArUS program, BAA number IARPA-BAA-10-04, via contract 2009-0917826-016, and is subject to the Rights in Data-General Clause 52.227-14, Alt. IV (DEC 2007). Any views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein.

inadequacies in the initial account. Comparison of alternative accounts. Reframing the initial account and replacing it with another. The deliberate construction of an account when none is automatically recognized."

But as described above and throughout Klein et al. (2007), it is not clear exactly what *entities* or *elements* are fitted into frames, or how *accounts* are formed from frames and used to explain *events*. Even the exact form of a *frame* is not clear, as Klein et al. (2007) say: *"A frame can take the form of a story... map... script... plan... [or other] structure for accounting for the data and guiding the search for more data."*

A Bayesian Approach to Sensemaking

The representational structures and computational processes of sensemaking can be specified formally using Bayesian principles (Bayes, 1763; Fischhoff & Beyth-Marom, 1983; Mueller, 2009). Here the approach (Burns, 2005, 2011, 2014a) involves five distinct concepts collectively dubbed HELP: *hypotheses*, *evidence*, *likelihoods*, *priors*, and *posteriors*. The hypotheses are possible explanations of actual evidence that has been received or potential evidence that might be received. The likelihoods, priors, and posteriors are each represented by a probability ranging from zero to one. A likelihood, denoted $P(e|H)$, is the probability of some evidence (e) assuming the truth of a hypothesis (H). A prior, denoted $P(H)$, is the probability of a hypothesis in the absence of some evidence. A posterior, denoted $P(H|e)$, is the probability of a hypothesis (H) given some evidence (e).

The primary process of Bayesian inference is one of updating prior probabilities to compute posterior probabilities, and this applies iteratively. That is, the posterior probability of a hypothesis (after some evidence) becomes the prior probability for that hypothesis in a future update with further evidence. The updating is accomplished using Bayes' Rule, which states that a posterior probability is computed as the normalized product of a prior and likelihood, $P(H_i|e) = P(H_i) * P(e|H_i) / P(e)$. The normalizing factor $P(e)$ is a marginal probability computed as the sum of products $P(H_i) * P(e|H_i)$ over all hypotheses in a set $\{H_i\}$ of mutually exclusive and exhaustive hypotheses.

These concepts of Bayesian HELP serve to formalize the notion of a frame, by defining a frame as a knowledge structure comprising hypotheses, evidence, likelihoods, priors, and posteriors. Unlike the data-frame distinction between data and frame, a Bayesian frame includes data (i.e., evidence) as well as other knowledge and beliefs (i.e., hypotheses, likelihoods, priors, and posteriors) by which one makes sense of the data. The reason is that likelihoods are needed for computing confidence in hypotheses, and likelihoods always refer to data (evidence) because a likelihood is the probability of some evidence given a hypothesis.

The concepts of Bayesian HELP also serve to formalize the notions of framing and reframing, i.e., as processes for computing confidence across a set of hypotheses. In fact there are at least three different types of reframing that can be distinguished as follows: *updating*, *revising*, and *abducting*. In *updating* (described above), new evidence and associated likelihoods are used to update priors and compute posteriors via Bayes' Rule over a fixed set of hypotheses. In *revising*, old likelihoods are replaced by new likelihoods and a previous update is repeated, again over a fixed set of hypotheses. In *abducting*, new hypotheses are generated along with associated priors and likelihoods of evidence, and posteriors are computed over the new set of hypotheses.

A Real-World Story of Sensemaking

At this point some readers may be skeptical of a Bayesian approach to sensemaking. For example, one might argue that humans are not perfect Bayesians because sensemaking involves well-known heuristics and biases (Kahneman, 2011). But actually this is an advantage of the Bayesian approach, as leveraged in research on ICArUS (Burns, 2014a, 2014b; Burns, Greenwald & Fine, 2014), because it enables the modeling and measuring of heuristics and biases relative to normative standards.

On the other hand, one might argue that mathematical approaches cannot possibly capture the richness of psychological processes. But a bounded-Bayesian approach has already been used to analyse real-world command and control (Burns, 2005), and the same approach has even been used to compute the aesthetics of creative artworks (Burns, 2012, 2014c) for which sensemaking is arguably even more resistant to quantification. As shown in those studies, numbers are needed to apply the approach, but even rough estimates are sufficient to obtain results that are consistent with the qualitative judgments of humans.

Thus encouraged by these earlier efforts, Bayesian HELP is applied below to a real-world story of sensemaking by an intelligence analyst.

RESULTS

Klein et al. (2007) tell a true story involving five cycles of sensemaking, each addressed in a numbered subsection below.

1. Suspecting "The Bad Guys"

"Major A. S. discussed an incident that occurred soon after 9/11 in which he was able to determine the nature of overflight activity around nuclear power plants and weapons facilities. This incident occurred while he was

an analyst. He noticed that there had been increased reports in counterintelligence outlets of overflight incidents around nuclear power plants and weapons facilities. At that time, all nuclear power plants and weapons facilities were 'temporary restricted flight' zones. So this meant there were suddenly a number of reports of small, low-flying planes around these facilities. At face value it appeared that this constituted a terrorist threat—that 'bad guys' had suddenly increased their surveillance activities. There had not been any reports of this activity prior to 9/11 (but there had been no temporary flight restrictions before 9/11 either)."

This first cycle of sensemaking begins as the sensemaker (hereafter denoted M) attends to an item of evidence from counterintelligence, denoted here as *s = sudden increase (after 9/11) in reported flight zone violations*. M thought this constituted a terrorist threat, so he was generating hypotheses $\{H_i\}$ about possible causes of the evidence s and estimating likelihoods of the form $P(s|H_i)$. In fact mental likelihoods of the form $P(s|H_i)$ would govern which hypotheses are recalled or constructed from long-term memory and represented in working memory as possible explanations of the observed evidence s. The story mentions a hypothesis denoted here as *A = Al Qaeda*, and suggests there was a strong association between A and s in the mind of M such that P(s|A) was large. Although the story does not say, M would also have generated the hypothesis *~A = Not Al Qaeda*, to represent other possible explanations, because he was clearly not certain that the evidence s was caused by A. Finally, besides a set of at least two hypotheses {A, ~A}, and associated likelihoods P(s|A) and P(s|~A), M would also be representing prior probabilities P(A) and P(~A) in his working memory. These priors reflect preconceived beliefs that M brings to the first cycle of sensemaking without regard for the evidence s.

The story does not provide numerical values for any probabilities, and if asked the sensemaker M might even deny that he represented such quantities in his mind. But clearly M is not equally confident in A and ~A, so some measure of relative confidence in these two hypotheses is mentally represented at least implicitly. Similarly, likelihoods of the form P(s|A) and P(s|~A) are represented, at least implicitly, because these likelihoods govern which hypotheses are generated in the first place. For example, the story suggests that P(s|A) is much higher than P(s|~A), because M can think of a reason (i.e., surveillance by terrorists) why A would cause s but does not think of a reason why ~A would cause s.

The point here is twofold: First, hypotheses, evidence, likelihoods, priors, and posteriors (HELP) may all be represented in the mind of a sensemaker, at least implicitly and qualitatively, in order for the sensemaker to make sense of what has been sensed (as evidence). Second, the same components of HELP must be represented explicitly and quantitatively in order to rigorously model and measure sensemaking. Therefore, for purposes of quantification here, we can assign numbers that are at least roughly consistent with the story. For example, we might assume P(A) = P(~A) = 0.50 if M's prior confidence was indifferent between A and ~A. However, the events of the story took place soon after the 9/11 attacks when Al Qaeda was prominent in the thoughts of most Americans, so here as rough estimates we might assume P(A) = 0.80 and P(~A) = 0.20. Note that P(A) + P(~A) = 1, because A and ~A are mutually exclusive and exhaustive hypotheses.

Also consistent with the story, we might assume P(s|A) = 0.90 and P(s|~A) = 0.50 for the likelihoods of observing the evidence s if A or ~A were true, respectively. But notice that, unlike the priors, these likelihoods need not and usually will not sum to 1. Instead P(s|A) + P(~s|A) = 1, because if A is true then either s or ~s would occur. Thus the assumed value P(s|A) = 0.90 and corresponding value P(~s|A) = 1 - 0.90 = 0.10 together mean that M thinks Al Qaeda is much more likely to cause s than ~s, because M can think of a reason why A would cause s rather than ~s. Similarly, P(s|~A) + P(~s|~A) = 1, because if ~A is true then either s or ~s would occur. Here the assumed value P(s|~A) = 0.50 means that s would be a random (i.e., for no apparent reason) effect if ~A was true, such that P(s|~A) = P(~s|~A) = 0.50.

Using the priors and likelihoods outlined above, we can complete our Bayesian analysis of how the sensemaker formed his initial belief that s was most probably caused by "bad guys" (A). The posterior is computed as a normalized product of prior and likelihood, for each hypothesis (A and ~A), via Bayes' Rule as follows: P(A|s) = P(A) * P(s|A) / P(s); P(~A|s) = P(~A) * P(s|~A) / P(s), where P(s) is a normalizing factor appearing in the denominators, computed from the sum of numerators as follows: P(s) = P(A) * P(s|A) + P(~A) * P(s|~A). Using the numbers noted above, these equations produce posterior probabilities of P(A|s) = 0.88 and P(~A|s) = 0.12. In words, M would be thinking that Al Qaeda's surveillance activities are the most probable explanation of the evidence from counterintelligence.

2. Reviewing Their Tactics

"Major A. S. obtained access to the Al Qaeda tactics manual, which instructed Al Qaeda members not to bring attention to themselves. This piece of information helped him to begin to form the hypothesis that these incidents were bogus—'It was a gut feeling, it just didn't sit right. If I was a terrorist I wouldn't be doing this.' He recalled thinking to himself, 'If I was trying to do surveillance how would I do it?' From the Al Qaeda manual, he knew they wouldn't break the rules, which to him meant that they wouldn't break any of the flight rules. He asked himself, 'If I'm a terrorist doing surveillance on a potential target, how do I act?' He couldn't put

together a sensible story that had a terrorist doing anything as blatant as overflights in an air traffic restricted area."

Based on his posterior beliefs after assessing the evidence s, M would have formed expectations about further information that might be obtained and assessed next. Those expectations would affect whether he would seek more information (or not), and where he would seek to obtain it. The story tells us that M obtained access to the Al Qaeda manual, so apparently he expected it would say something that would shed light on the likelihood P(s|A). Although the story does not say, it is reasonable (Burns, 2005) to assume that M expected the manual would provide some information that confirms his suspicions about A, simply because at this point A was the most probable hypothesis. In that light M must have been *surprised* by what he read, because it was a violation of his expectations (Burns, 2012, 2014c). More specifically, M learned that Al Qaeda members are instructed not to bring attention to themselves, and this affected his estimate of the likelihood P(s|A).

For example, we might assume that after reading the Al Qaeda manual M thought P(s|A) = 0.01. In effect M realized that his previous estimate of P(s|A) = 0.90 was wrong, because he learned of a very good reason for why A would not cause s and instead would cause ~s. So M repeats the previous cycle of sensemaking, but now using P(s|A) = 0.01 instead of P(s|A) = 0.90. The Al Qaeda manual says nothing about other groups (~A), so P(s|~A) remains = 0.50.

Using the revised likelihoods, along with the original priors of P(A) = 0.80 and P(~A) = 0.20, the Bayesian equations produce posteriors as follows: P(A|s) = 0.07 and P(~A|s) = 0.93. In words, the sensemaker's beliefs have undergone a reversal, from A being very probable to ~A being very probable, based on a change in the likelihood P(s|A). So here we find a form of reframing that involves *revising* likelihoods and associated posteriors across a set of hypotheses {A, ~A}. This *revising* is the first of three fundamentally different types of reframing that are found in the story, and the other two types will be highlighted later when they occur.

As a result of revising likelihoods and posteriors, the story says that M *"began to form the hypothesis that these incidents were bogus"*. But notice that this is not really a new hypothesis, because the hypothesis ~A had been generated earlier along with the hypothesis A. Instead at this point M began to wonder who, if not Al Qaeda, is likely to break the rules and cause the observed evidence s. Eventually M generated a new hypothesis in answer to this question, but it was not until the next cycle of sensemaking. What is interesting here in the present cycle is that M felt compelled to think deeper about the hypothesis ~A, in light of the evidence s. In doing so it appears that M was motivated by two things. First, he now thought ~A was the most probable hypothesis. Second, his likelihoods for this most probable hypothesis were P(s|~A) = 0.50 and P(~s|~A) = 0.50, so M had no causal basis or reason by which he could explain the evidence s. In other words, M was pretty sure he knew who was <u>not</u> responsible for the overflight activity, but he still had no clue as to who <u>was</u> responsible – and apparently he felt a strong need to establish who was responsible.

3. Abducting a Reason

"He thought about who might do that, and kept coming back to the overflights as some sort of mistake or blunder. That suggested student pilots to him because 'basically, they are idiots.' He was an experienced pilot. He knew that during training, it was absolutely standard for pilots to be instructed that if they got lost, the first thing they should look for were nuclear power plants. He told us that 'an entire generation of pilots' had been given this specific instruction when learning to fly. Because they are so easily sighted, and are easily recognized landmarks, nuclear power plants are very useful for getting one's bearings. He also knew that during pilot training the visual flight rules would instruct students to fly east to west and low—about 1,500 feet. Basically students would fly low patterns, from east to west, from airport to airport."

Motivated by his desire to find a causal reason for the evidence s, M initiated this third cycle of sensemaking without the introduction of any new information. That is, M was generating hypotheses about who might be responsible for s, after realizing that Al Qaeda (A) is probably not responsible.

The result is a new hypothesis *S = Student pilots (and not Al Qaeda)*, based on a strong association between S and s in M's mind, which reflects a reason for why S would cause s. That is, based on M's expertise as a pilot, he thinks P(s|S) is high because he knows why students would be likely to fly over nuclear power plants. Numerically, we might assume P(s|S) = 0.90 because students have a reason for causing s, and P(s|~S) = 0.50 because non-students may or may not have a reason for causing s.

At this point M's set of hypotheses can be characterized as {A, S, ~S}, where *~S = Not student pilots (and not Al Qaeda)*. Also at this point M's reframing involves *abducting* hypotheses and associated likelihoods of those hypotheses. This is much like the initial framing we saw in the first cycle of sensemaking, and it is clearly more complex than the *revising* (over a fixed set of hypotheses) that we saw in the second cycle.

To complete the analysis of this third cycle, we can assume P(A) = 0.80 as before, and then assume P(~A) = 0.20 is split equally between the two hypotheses that were not previously distinguished within ~A such that P(S) = P(~S) = 0.10. For likelihoods, we have P(s|A) = 0.01 from the previous cycle of sensemaking, and now from the present cycle we have P(s|S) = 0.90 and P(s|~S) = 0.50. Using Bayes' Rule to compute the posteriors yields:

P(A|s) = 0.05, P(S|s) = 0.61, and P(~S|s) = 0.34. In words, M thinks S is about ten times more probable than A, and he also thinks S is about twice as probable as ~S.

4. Collecting More Data

"It took Major A. S. about 3 weeks to do his assessment. He found all relevant message traffic by searching databases for about 3 days. He picked the three geographic areas with the highest number of reports and focused on those. He developed overlays to show where airports were located and the different flight routes between them. In all three cases, the 'temporary restricted flight' zones (and the nuclear power plants) happened to fall along a vector with an airport on either end. This added support to his hypothesis that the overflights were student pilots, lost and using the nuclear power plants to reorient, just as they had been told to do."

As in the second cycle of sensemaking, where M thought to consult the Al Qaeda manual, his beliefs here at the start of the fourth cycle led him to seek further information that might better distinguish the cause (A, S, or ~S) of evidence s. The story does not say why M chose to examine flight paths. But like his earlier decision to read the Al Qaeda manual, it is reasonable (Burns, 2005) to assume that he expected a flight path analysis would confirm his suspicions about the most likely hypothesis (S).

M's assessment of flight paths was a form of "suitability analysis", which is typically performed by geospatial analysts to establish whether features of terrain are likely to be suitable for some hypothesized activity. In this case M found that vectors through restricted zones had airports on either end, and the story says this added support to his hypothesis (S). But actually M's findings first affected his estimates of likelihoods, which in turn affected his posterior confidence in each hypothesis {A, S, ~S}. More specifically, M's finding that some vectors between airports passed directly over nuclear power plants led him to increase the likelihood P(s|S) and decrease the likelihood P(s|~S), relative to his earlier estimates for these same likelihoods. In that respect the reframing here is a *revising* of likelihoods and associated posteriors, similar to the *revising* we saw in the second cycle where M decreased his estimate for P(s|A) after reading the Al Qaeda manual.

For example, based on his geospatial analysis, we might assume M increased P(s|S) from 0.90 to 0.95 and decreased P(s|~S) from 0.50 to 0.10. The increase in P(s|S) reflects M's finding of airport vectors over nuclear plants, which make these paths quite suitable for lost students. The decrease in P(s|~S) comes from the finding of other flight paths that would be more suitable for experienced pilots.

Assuming the revised likelihoods are P(s|A) = 0.01, P(s|S) = 0.95, P(s|~S) = 0.10, and using the previous cycle's priors of P(A) = 0.80, P(S) = 0.10, and P(~S) = 0.10, the Bayesian posteriors are computed are follows: P(A|s) = 0.07, P(S|s) = 0.84, and P(~S|s) = 0.09. In words, M now thinks that S is about ten times more probable than either A or ~S, and M is even more certain than before that the most probable explanation for the overflight activity is student pilots (who are not members of Al Qaeda).

5. Concluding "It's Students"

"He also checked to see if any of the pilots of the flights that had been cited over nuclear plants or weapons facilities were interviewed by the FBI. In the message traffic, he discovered that about 10% to 15% of these pilots had been detained, but none had panned out as being 'nefarious pilots'. With this information, Major A. S. settled on an answer to his question about who would break the rules: student pilots. The students were probably following visual flight rules, not any sort of flight plan. That is, they were flying by looking out the window and navigating."

An interesting aspect of this story is that M chose to spend days or weeks on the flight path analysis, which would only help distinguish S from ~S, before checking the FBI records. The FBI records would help distinguish A from ~A, and a threat of Al Qaeda activity was M's primary concern at the start of the story. But here it appears that M's priority for further analysis was to establish who <u>did</u> cause s (which he suspected was S) rather than who did <u>not</u> cause s. Some might characterize this behavior as a confirmation bias (Nickerson, 1998), because M first chose to collect evidence that pertains to a more probable (and less consequential) hypothesis S, rather than collect evidence that pertains to a less probable (and more consequential) hypothesis A. But in fact M's behavior may actually be optimal, because a "positive test strategy" (Klayman & Ha, 1987) has been shown to maximize the expected gain in information for prototypical situations of intelligence collection (Burns, 2014a; 2014b). Also, if M's objective was to recommend some policy action to mitigate flight zone violations, then he would want and need to know who are the culprits rather than who are not the culprits. Thus like the earlier instances where M chose to obtain evidence that he expected would support his favored hypothesis, it is not clear whether M's confirmation preference is actually a confirmation bias (relative to Bayesian standards). An answer to that question would require that more parameters of the situation be identified and quantified.

In any case, the new evidence obtained in this fifth and final cycle of sensemaking is: *n = no nefarious pilots identified in the FBI interviews*. The associated likelihoods are probabilities of n conditional on each hypothesis {A, S, ~S}, but also conditional on the previous evidence s. Because n comes from a different and diverse

source of intelligence than the evidence s from counterintelligence, we can assume n and s are independent such that the likelihoods of n are conditioned only on hypotheses as follows: P(n|A), P(n|S), and P(n|~S). For example, based on the sample of pilots that had been interviewed, a finding of no nefarious pilots might suggest P(n|A) = 0. But because the sample is limited to 10-15% of pilots, and because interviews of pilots would not be 100% reliable in establishing ties to Al Qaeda, we might assume P(n|A) = 0.01 and P(~n|A) = 0.99. On the other hand, it appears the FBI data were uninformative with respect to the student status of pilots. So for students we have P(n|S) = P(~n|S) = 0.50, and also for non-students we have P(n|~S) = P(~n|~S) = 0.50.
Thus the three likelihoods for n are: P(n|A) = 0.01, P(n|S) = 0.50, and P(n|~S) = 0.50, and Bayes' Rule is used to update the posteriors computed in the previous cycle of sensemaking. Those posteriors become priors in the present cycle as follows: P(A|s) = 0.07, P(S|s) = 0.84, and P(~S|s) = 0.09. Combining these priors with the likelihoods via Bayes' Rule we obtain the following posteriors: P(A|n,s) = 0.001, P(S|n,s) = 0.90, and P(~S|n,s) = 0.10. In words, after five cycles of sensemaking the sensemaker M is now very sure the evidence (s and n) is not explained by Al Qaeda activity, P(A|n,s) = 0.001. He is also pretty sure that the evidence is explained by activities of student pilots following visual flight rules, P(S|n,s) = 0.90.
Notice the nature of reframing here in this final cycle is one of *updating* confidence in hypotheses, over a fixed set of hypotheses, based on likelihoods of the new evidence. This *updating* is different from the *abducting* we saw in framing and reframing during the first and third cycles, respectively, because here no new hypotheses are generated. This *updating* is also different from the *revising* we saw in the second and fourth cycles of sensemaking, because here the new likelihoods are used to augment previous likelihoods in an iterative Bayesian update, rather than to replace old likelihoods and repeat an old update. In iterative updating, posteriors from the previous update become priors for the present update.

DISCUSSION

The Nature of Reframing

As exposed in the above analysis, there are three fundamentally different types of *"reframing"* that are made explicit by Bayesian HELP, namely: *updating*, *revising*, and *abducting*. All three types were carefully considered in the design of challenge problems for IARPA's program ICArUS (Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking), to ensure that experiments were as naturalistic as possible.
Because of program constraints (IARPA, 2010), the challenge problems were designed primarily to address *updating*, and secondarily to address *revising*. One constraint was to minimize the role of rich background knowledge possessed by human participants, because it was infeasible to provide neural-computational models with the same background knowledge – as would be needed to control for expertise when comparing human and model performance in sensemaking experiments. But as seen in the story analysed above, expert knowledge is how humans generate hypotheses (and estimate likelihoods, and establish priors). So the upshot is that ICArUS challenge problems involved only fixed sets of hypotheses, which were provided to participants, and experiments could not address *abducting* as a form of reframing. Another constraint was that each ICArUS experiment was to measure and model average human performance across N ≈ 100 participants. As a practical matter, this required that all participants use the same evidence and likelihoods (as well as the same hypotheses), rather than allowing each participant to collect their own evidence and estimate their own likelihoods (which would have been more akin to ≈ 100 experiments, each with N = 1). These constraints on evidence and likelihoods were loosened in some trials of the experiments, in order to study human decisions made in collecting evidence and human inferences made in *revising* likelihoods and posteriors. But the bulk of trials focused on *updating* confidence across hypotheses, using evidence and likelihoods that were provided to all participants as inputs to sensemaking.

An Insight on Biases

Although ICArUS challenge problems were not completely naturalistic with respect to *abducting*, the experiments offer useful insights into *revising* and especially *updating*. In particular, results showed that humans were biased in their Bayesian updates because they were "substituting" simple heuristics (Kahneman, 2011) for the more complex calculations of Bayes' Rule. The most common error was to compute a posterior as the average of a prior and likelihood, rather than the normalized product of a prior and likelihood per Bayes' Rule. The resulting posteriors were conservative (Edwards, 1982), i.e., too close to {0.50, 0.50} compared to the Bayesian posteriors for the case of two hypotheses {H, ~H}, which means humans did not extract all the certainty that was available in the information they were given. Of course participants did not know they were substituting the wrong strategy, or else they would not have done so. So the obvious way to help humans overcome conservatism and other biases is simply to teach them the structure of Bayesian inference in the first place (Burns, 2006, 2007).

Toward that end, the present paper offers two contributions that might be used to improve intelligence analysis. The first contribution is formalizing the principles of sensemaking in Bayesian HELP. The second contribution is demonstrating how the principles of Bayesian HELP can be applied to a real-world story of sensemaking.

HELP Technique and HELP Training

Currently there exist numerous Structured Analytic Techniques (SATs) intended to aid intelligence analysts (Beebe & Pherson, 2012). However none of these SATs provides the requisite structure to support reasoning in accordance with Bayesian principles. The one SAT that comes closest (Heuer, 1999) was developed to help analysts overcome confirmation bias and is called Analysis of Competing Hypotheses (ACH). But as detailed elsewhere (Burns, 2014a), ACH does not address four classes of errors (and may even magnify such errors) commonly found in biased inferences (Burns, 2006, 2007), namely: (1) failure to generate a mutually exclusive and exhaustive set of *hypotheses*, (2) failure to distinguish assumptions from *evidence*, (3) failure to distinguish *likelihoods* from posteriors, and (4) failure to properly aggregate priors and likelihoods in computing *posteriors* (e.g., the "averaging" heuristic that leads to the conservative bias discussed above).

All four classes of errors are addressed by Bayesian HELP (Burns, 2014a), which suggests that this structure could be used to support and improve intelligence analysis. However no SAT is of benefit unless it can be learned and applied in practice. The present paper illustrates how HELP can be taught, using stories to engage analysts in relevant case studies that could be tailored to their interests and expertise. Unlike laboratory experiments, the story analysed here includes all the richness of naturalistic sensemaking to which HELP applies. Other stories could be analysed in a similar fashion, as examples that are provided to students or as exercises to be performed by students. In most cases, including the story analysed here, there may not be a single correct solution. But developing and debating possible solutions, using the structured technique of Bayesian HELP, could improve the rigor of analytic reasoning under uncertainty.

CONCLUSION

This paper demonstrated how Bayesian HELP (hypotheses, evidence, likelihoods, priors, and posteriors) can formalize notions of *"frames"*, *"framing"*, and *"reframing"* that appear in theories of sensemaking. HELP has been used to dissect real-world intelligence and design research experiments. HELP can also be used as a structured technique for improving naturalistic sensemaking in the field of intelligence analysis.

REFERENCES

Bayes, T. (1763). An essay toward solving a problem in the doctrine of chances. *Philosophical Transactions, 53*, 370-418.

Beebe, S. & Pherson, R. (2012). *Cases in Intelligence Analysis: Structured Analytic Techniques in Action*. Los Angeles, CA: Sage CQ Press.

Burns, K. (2005). Mental models and normal errors. In H. Montgomery, R. Lipshitz & B. Brehmer (Eds.), *How Professionals Make Decisions* (pp. 15-28). Mahwah, NJ: Lawrence Erlbaum.

Burns, K. (2006). Bayesian inference in disputed authorship: A case study of cognitive errors and a new system for decision support. *Information Sciences, 176*, 1570-1589.

Burns, K. (2007). Dealing with probabilities: On improving inferences with Bayesian Boxes. In R. Hoffman (Ed.), *Expertise Out of Context* (pp. 263-280). New York: Lawrence Erlbaum.

Burns, K. (2011). The challenge of iSPIED: Intelligence sensemaking to prognosticate IEDs. *The International Command and Control Journal, 5(1)*, 1-38.

Burns, K. (2012). EVE's energy in aesthetic experience: A Bayesian basis for haiku humor. *Journal of Mathematics and the Arts, 6*, 77-87.

Burns, K. (2014a). ICArUS: A Computational Basis for ICArUS Challenge Problem Design. *MITRE Technical Report, MTR 149415.*

Burns, K. (2014b). ICArUS: Phase 2 Challenge Problem Design and Test Specification. *MITRE Technical Report, MTR 149412.*

Burns, K. (2014c). Computing the creativeness of amusing advertisements: A Bayesian model of Burma-Shave's muse. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, http://dx.doi.org/10.1017/S0890060414000699.

Burns, K., Bonaceto, C., Fine, M. & Oertel, C. (2014). ICArUS: Overview of Test and Evaluation Materials. *MITRE Technical Report, MTR 149409*.

Burns, K., Greenwald, H. & Fine, M. (2014). ICArUS: Phase 1 Challenge Problem Design and Test Specification. *MITRE Technical Report, MTR 149410*.

Edwards, W. (1982). Conservatism in human information processing. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment Under Uncertainty: Heuristics and Biases* (pp. 359-369). Cambridge: Cambridge University Press.

Fischhoff, B. & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review, 90(3)*, 239-260.

Heuer, R. (1999). *Psychology of Intelligence Analysis*. Washington, DC: Center for the Study of Intelligence, Central Intelligence Agency.

IARPA (2010). Broad Agency Announcement, Integrated Cognitive-neuroscience Architectures for Understanding Sensemaking (ICArUS). *Intelligence Advanced Research Projects Activity (IARPA), IARPA-BAA-10-04*. April 1, 2010.

Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Strauss & Giroux.

Klayman, J. & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94(2)*, 211-228.

Klein, G., Moon, B. & Hoffman, R. (2006a). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems, 21(4)*, 70-73.

Klein, G., Moon, B. & Hoffman, R. (2006b). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems, 21(5)*, 88-92.

Klein, G., Phillips, J., Rall, E. & Peluso, D. (2007). A data-frame theory of sensemaking. In R. Hoffman (Ed.), *Expertise Out of Context* (pp. 113-155). New York: Lawrence Erlbaum.

Mueller, S. (2009). A Bayesian recognitional decision model. *Journal of Cognitive Engineering and Decision Making, 3(2)*, 111-130.

Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2(2)*, 1

Decision Support in Team Collaboration for Cyber Situation Awareness

Zequn Huang[a] Chien-Chung Shen[a] Sheetal Doshi[b]
Nimmi Thomas[b] Ha Duong[b] Cliff Wang[c]
[a]Department of Computer & Information Sciences
University of Delaware, Newark DE 19716
{zequn,cshen@udel.edu}
[b]Scalable Network Technologies, Inc.
[c]Army Research Office

ABSTRACT

Complex and dynamically changing tasks such as cyber defence often require the effective coordination of a team of cyber analysts that work at different levels and/or different parts of the system. Each team member collects data, generates his/her own awareness for the current situation, and shares the awareness with other members to generate the comprehensive understanding of the overall situation for the purpose of decision-making. Since each team member may have his/her own personal expertise knowledge, experience, and opinions, it is difficult for the whole team to make consensus decision when having conflicting judgments on the cyber situation. Considering human cyber analysts tend to use ambiguous linguistic language to express their own cyber situation awareness during team discussion, we propose a fuzzy logic based method to facilitate cyber analysts to quantify their preference and make consensus decision that is most acceptable by the entire team.

KEYWORDS

Cyber Situation Awareness; Team Collaboration; Fuzzy Logic; Multi-criteria Team Decision Making.

INTRODUCTION

Cyber Situation Awareness (CSA) supports decision-making and responses of cyber analysts by understanding the overall context of network vulnerabilities, how they are interrelated, and how attacks may exploit them to penetrate deeper in the network. When confronted with sheer amount of situation information and dynamically changing environments, cyber analysts, at different levels and/or in different parts of the system, need to work collaboratively as a team. Typically, each team member forms his/her own CSA and shares it with other team member in order to create *team CSA*. However, each individual cyber analyst may have his/her own personal expertise, experience, and opinions, so that conflicts often occur in team decision-making on, for instance, whether there exists a cyber attack, false alarms, as well as the types of detected cyber attacks. Therefore, how to resolve conflicts within team CSA becomes a critical issue. Traditional methods for a team of cyber analysts to achieve consensus decision, such as verbal discussion and whiteboard session, are not accurate and unpersuasive. In this paper, we investigate a fuzzy logic based approach that can aggregate uncertain information to generate consensus CSA for the entire team.

BACKGROUND AND RELATED WORK

According to the reference model proposed by Endsley (Sushil & Peng, 2010), Situation Awareness (SA) is a three phases process: *perception*, *comprehension*, and *projection*. SA begins with perception. Perception provides information about the status, attributes, and dynamics of relevant elements within the environment. It also includes classifying information into understood representations and provides the basic building blocks for comprehension and projection. Comprehension of the situation encompasses how people combine, interpret, and correlate information. Thus, comprehension includes more than perceiving or attending to information; it includes the integration of multiple pieces of information and a determination of their relevance. Comprehension yields an organized picture of the current situation by determining the significance of objects and events. Furthermore, as a dynamic process, comprehension must combine new information with already existing knowledge to produce a composite picture of the situation as it evolves in the future, which is projection. Cyber Situation Awareness is SA extended to the cyber domain. Similarly, CSA is also a three phases process: collect data and seek cues that form attack tracks; estimate impact of observed attack tracks; anticipate moves (actions, targets, time) of attackers.

A team is defined as a group of 'heterogeneous' people working together towards a common goal. The heterogeneity could be based on their individual skill, information they know, or the resources they have. Team situation awareness is defined as the degree to which every team member possesses the SA required for his or her responsibilities. The team members through team interactions transform individual knowledge to collective knowledge to achieve team situation awareness (Michael & Prashanth, 2012; Nancy & Michael, 2013). However, team situation awareness is more than the sum of situation awareness of the individuals in the team. Normally, in a team, each member holds his/her own component of SA and share information with other team members. Since each team member may have his/her own personal expertise knowledge, experience, and opinions, it may be difficult for them to make consensus decision. Besides, cyber analysts often describe situations with imprecisely or ambiguously information, that exacerbates the uncertainty of shared situation awareness.

CyberCog (Prashanth, 2011) is a synthetic task environment for understanding and measuring individual and team situation awareness, and for evaluating algorithms and visualization intended to improve cyber situation awareness. CyberCog provides an interactive environment for conducting human-in-the-loop experiment in which the participants of the experiment perform the tasks of cyber analysts in response to cyber attack scenarios. CyberCog generates performance measures and interaction logs for measuring individual and team performance. CyberCog utilizes a collection of known cyber defence incidents and analysis data to build a synthetic task environment. Alerts and cues are generated based on emulation of real-world analyst knowledge. From the mix of alerts and cues, cyber analysts will react to identify threats and vulnerabilities individually or as a team. The identification of attacks is based on knowledge about the attack alert patterns.

In the scenarios developed by CyberCog, cyber analysts can work together as a team. For instance, each cyber analyst receives individualized training on his/her specific role, such as Malware specialist, Denial of Service specialist and Phishing attack specialist. During training, if one cyber analyst encounters alerts that he/she is not very familiar with, he/she can share the alerts with the rest of the team to ask other cyber analysts for help. CyberCog provides a collaboration tool called Shared Events Viewer, through which team members could share event information to get help with unfamiliar event patterns. Other team members may reply to a shared event with details and information on what needs to be done and how to carry out an investigation process for this event pattern. This interaction is very similar to interaction patterns among cyber analysts in the real world. However, one major limitation of CyberCog team CSA is the lack of a quantification method to aggregate individual CSA to generate team consensus.

DECISION SUPPORT IN TEAM COLLABORATION FOR CYBER SITUATION AWARENESS

The primary role of human cyber analysts includes: collecting and filtering computer network traffic, the traffic for suspicious or unexpected behaviour, and discovering system misuse or unauthorized system access. The cyber analyst's transform observed alerts into their own cyber situation awareness with descriptions such as follows:

- High memory usage on host A
- Very low CPU utilization rate on host B
- Unusually large date uploads from host C
- Excessive failed login's from a remote source IP
- Host D receives UDP packets with extremely large payloads

Notice that human cyber analysts often know a situation imprecisely or ambiguously and use non-quantitative qualifiers, such as 'excessive', 'high' and 'very low' to describe such situation. Furthermore, team members may have conflicting judgements (individual CSA) about the type of the current attack due to their own personal expertise knowledge and experience. As team CSA has to be generated through aggregating these imprecise and inaccuracy information, aggregating individual cyber situation awareness and making consensus decision becomes a critical issue.

In this paper, we propose to utilize *fuzzy logic* to let team members achieve consensus awareness for the situation. Fuzzy logic (Kwang 2004) has been well applied in the area of multi-criteria team decision-making to deal with uncertain issues in generating a consensus opinion, such as facility location selection (Fatih & Serkan, 2009; Herrera & Verdegay, 1996; Cengiz & Da, 2003). It can construct preference relation between alternatives by evaluating different criteria, and select the best action from a set of alternatives that is most acceptable by the entire team. In particular, we use fuzzy logic to facilitate a team of cyber analysts to make consensus decision when they have conflicting judgements on the type of on-going attacks.

First, we present notations used. Let $P = \{P_1, P_2, \ldots, P_n\}$, $n \geq 2$, be a given finite set of decision makers, which contains at least two cyber analysts, to select a satisfactory solution from a set of alternatives; $S = \{S_1, S_2, S_3, \ldots, S_m\}$, $m \geq 3$, be a given finite set of alternatives for a decision problem; $C = \{C_1, C_2, \ldots, C_t\}$, $t \geq 2$ be a given finite set of selection criteria for the decision alternatives. The procedure of team decision-making consists of the following eight steps:

- Step 1: Determine solution alternatives
- Step 2: Choose criteria
- Step 3: Determine the weights of decision makers
- Step 4: Determine the weights of criteria
- Step 5: Construct belief level matrix
- Step 6: Construct the aggregated weighted team fuzzy decision matrix
- Step 7: Obtain fuzzy positive-ideal solution and fuzzy negative-ideal solution
- Step 8: Calculate the closeness coefficient and rank the solution alternatives

Step 1: Determine solution alternatives

When a decision problem of identifying the type of cyber attack is presented to a team, cyber analysts may propose different alternatives regarding the type of the on-going cyber attack. After combining all the possibilities, the alternatives set is defined as $S = \{S_1, S_2, S_3, \ldots, S_m\}$, $m \geq 3$. Figure 1 depicts an example GUI for team members to suggest possible alternative types of cyber attacks.

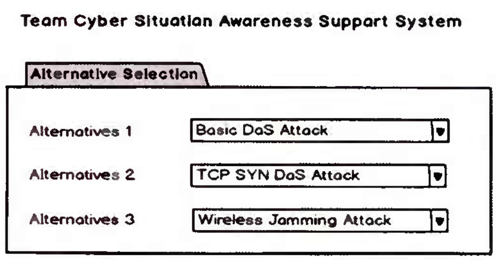


Figure 1. Three attack type alternatives proposed by team members

Step 2: Choose criteria

Each team member can also propose several criteria for assessing these alternatives. Criteria proposed by all the team members' are put into a criteria pool. If the criteria pool becomes too big, only the top-t criteria, $C = \{C_1, C_2, \ldots, C_t\}$, $t \geq 2$, can be chosen for the purpose of computational efficiency. The process of choosing criteria can be done via team discussion or voting among team members as shown in Figure 2.

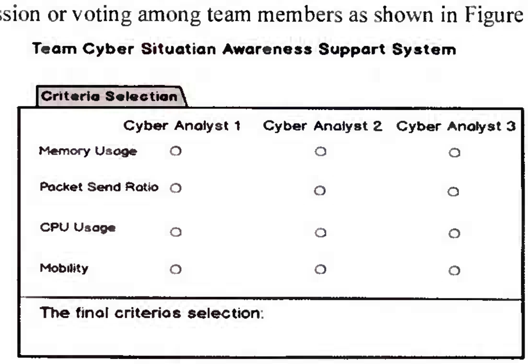


Figure 2. Solution criteria selection system

Step 3: Determine the weights of decision makers

As team members may have different degrees of influent on the decision selection, they should be assigned with different weights. Cyber analysts will be given an evaluation (test) and assigned weights based on the individual performance; the corresponding weights of decision makers are presented in Table 1.

Table 1. Individual cyber analyst performance

Individual Performance	Weight
Excellent	4.0
Great	3.7
Good	3.3
Normal	3.0
Not Good	2.0

Each member (decision maker) P_k ($k = 1, 2, ..., n$) is assigned weight that describe his/her influence on decision making. Then, the normalized weight vector is denoted as:

$$v = (v_1, v_2, \ldots, v_n) \text{ and } \sum_{k=1}^{n} v_k = 1$$

Step 4: Determine the weights of criteria

Each decision maker should determine the weight of selection criteria C_i through *pairwise comparison* with other criteria C_j. The linguistic terms of comparison are shown in Table 2. The comparison scale ranges from 1 to 9, representing the concepts of: 1 - equally important; 3 - weakly more important; 5 - strongly more important; 7 - demonstratively more important; 9 - absolutely more important. Values 2, 4, 6 and 8 are intermediate values between adjacent judgments.

Table 2. Linguistic terms for the comparison of criteria

Linguistic Terms	Comparison Scale
Equally important	1
Weakly more important	3
Strongly more important	5
Demonstratively more important	7
Absolutely more important	9

Using on the comparison scale, each team member fills the pairwise comparison in the GUI shown in Figure 3.

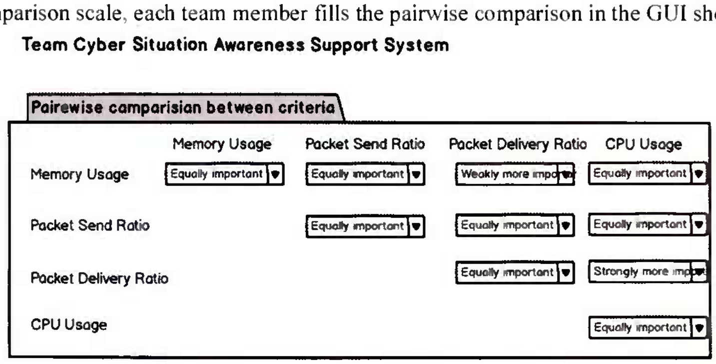


Figure 3. Criteria pairwise comparison

By pairwise-comparing the relative importance of selection criteria, the criteria comparison matrix $E^k = [e_{ij}]$ for decision maker P_k ($k = 1, 2, ..., n$) is generated. The example criteria comparison matrix from decision maker P_1 corresponding to his/her selection on Figure 3 is as follows:

$$E^1 = \begin{vmatrix} 1 & 1 & 3 & 1 \\ 1 & 1 & 1 & 1 \\ 1/3 & 1 & 1 & 5 \\ 1 & 1 & 1/5 & 1 \end{vmatrix}.$$

In order to generate consistent weights for every selection criterion, the *geometric mean* of each row of the criteria comparison matrix is calculated and then the results are normalized.

$$w^1 = \begin{pmatrix} (1 \times 1 \times 3 \times 1)^{1/4} \\ (1 \times 1 \times 1 \times 1)^{1/4} \\ (1/3 \times 1 \times 1 \times 5)^{1/4} \\ (1 \times 1 \times 1/5 \times 1)^{1/4} \end{pmatrix} \Rightarrow \begin{pmatrix} 1.31607 \\ 1.0 \\ 1.13622 \\ 0.66874 \end{pmatrix} \Rightarrow \begin{pmatrix} 1.31607/4.12103 \\ 1.0/4.12103 \\ 1.13622/4.12103 \\ 0.66874/4.12103 \end{pmatrix} \Rightarrow \begin{pmatrix} 0.3194 \\ 0.2427 \\ 0.2757 \\ 0.1622 \end{pmatrix}$$

The criteria weights for decision maker P_k is denoted as

$$w^k = (w_1^k, w_2^k, w_3^k, \ldots, w_t^k) \text{ and } \sum_{i=1}^{t} w_i = 1$$

Step 5: Construct the belief level matrix

Against every selection criterion C_i ($i = 1, 2, \ldots, t$), a belief level can be introduced to express the possibility of selecting solution S_j ($j = 1, 2, .., m$) under criterion C_i for decision maker P_k ($k = 1, 2, .., n$). The belief level matrix for decision maker P_k is denoted as b^{ij} ($i = 1,2,..,t; j = 1,2,..,m$) which belongs to a set of linguistic terms that contain various degrees of preferences required by decision maker P_k. The linguistic terms for variable preference are shown in Table 3.

Table 3. Linguistic terms for preference belief levels for alternatives

Linguistic Terms	Fuzzy numbers
Highest	0.36
High	0.28
Medium	0.20
Low	0.12
Lowest	0.04

Each cyber analyst fills up a belief level matrix to express his/her selection under the four selected criteria with three alternatives in the GUI depicted in Figure 4.

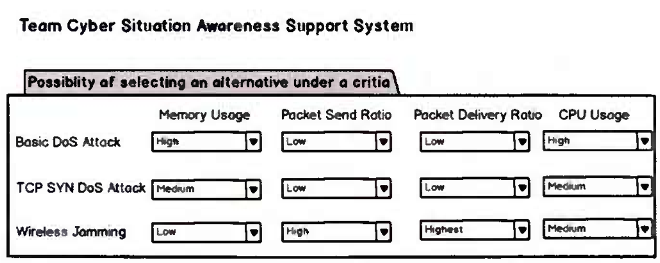


Figure 4. Belief level matrix filled by one cyber analyst

Elements in each belief level matrix b_{ij} is aggregated into belief vectors b_j by multiplying criteria weight vector, which stands for decision maker P_k's belief on the jth alternative as:

$$b_j^k = (w_1^k * b_{j1}^k + w_2^k * b_{j2}^k + \ldots + w_t^k * b_{jt}^k),\ j = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n$$

Step 6: Construct the aggregated weighted team fuzzy decision matrix

Based on the normalized decision maker weight vector and belief vector, we can construct a weighted fuzzy decision vector.

$$\left(r_1^{'}, r_2^{'}, \dots, r_m^{'}\right) = (v_1, v_2, \dots, v_n)\begin{pmatrix} b_1^1 & b_2^1 & \cdots & b_m^1 \\ b_1^2 & b_2^2 & \cdots & b_m^2 \\ \vdots & \vdots & \ddots & \vdots \\ b_1^n & b_2^n & \cdots & b_m^n \end{pmatrix}$$

The normalized decision vector is denoted as $r = \{r_1, r_2, \dots, r_m\}$ and each element r_j is calculated as:

$$r_j = \frac{r_j^{'}}{\sum_{i=1}^{n} r_j^{'}}, \text{ j} = 1, 2, \dots, \text{m}$$

Step 7: Obtain fuzzy positive-ideal solution and fuzzy negative-ideal solution

The basic principle is that the chosen alternative should have the shortest distance from the ideal solution and the farthest distance from the negative-ideal solution. In the weighted normalized fuzzy decision vector, r_j belongs to the close interval [0,1]. We can then define a fuzzy positive-ideal solution r^+ as 1 and a fuzzy negative-ideal solution r^- as 0. The positive and negative solutions whose distances between each r_j and r^+, and each r_j and r^- can be calculated as:

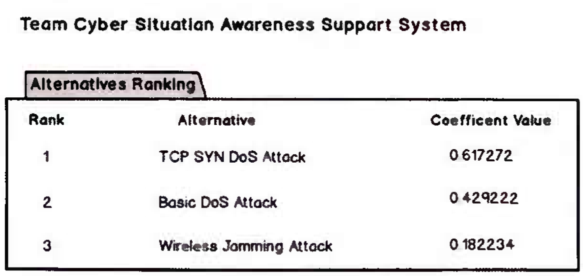
$$d_j^+ = d(r_j\ ,\ \mathrm{r}^+),\ \text{j} = 1, 2, \dots, \text{m}$$
$$d_j^- = d(r_j\ ,\ \mathrm{r}^-),\ \text{j} = 1, 2, \dots, \text{m}$$

Step 8: Calculate the closeness coefficient and rank the alternatives

A closeness coefficient is defined to determine the ranking order of all solutions once d^+ and d^- of each decision solution S_j ($j = 1, 2, .., m$) are obtained. The closeness coefficient of each solution is calculated as:

$$CC_j = (d_j^+ + (1 - d_j^-))/2,\ \text{j} = 1, 2, \dots, \text{m}$$

The alternative S_j that corresponds to $Max(CC_j)$ is the most acceptable solution for the decision team. As shown in Figure 5, based on the calculated coefficient values, the consensus decision in team CSA is that the current on-going attack is a TCP SYN DoS Attack.



Team Cyber Situation Awareness Support System

Alternatives Ranking

Rank	Alternative	Coefficent Value
1	TCP SYN DoS Attack	0.617272
2	Basic DoS Attack	0.429222
3	Wireless Jamming Attack	0.182234

Figure 5. Alternatives rank based on coefficient value

CONCLUSION

Due to the sheer amount of information generated by the cyber space, cyber analysts need to work collaboratively as a team at different levels and in different parts of the system. In this paper, we propose to use fuzzy logic to aggregate individual CSA into team CSA and make consensus decisions.

ACKNOWLEDGMENTS

This material is based upon work supported by US Army Research Office under contract W911NF-14-C-0140. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of US Army Research Office.

REFERENCES

Sushil Jajodia, Peng Liu, Vipin Swarup, and Cliff Wang (2010). Cyber Situational Awareness: Issues and Research. *Springer-Verlag New York Inc.* (pp. 3-13), 2010.

Michael A. Champion, Prashanth Rajivan, Nancy J. Cooke, and Shree Jari-wala (2012). Team-Based Cyber Defense Analysis. *Proceedings of 2012 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA 2012),* New Orleans, LA, USA, 2012.

Nancy J. Cooke, Michael Champion, Prashanth Rajivan, and Shree Jariwala (2013). Cyber Situation Awareness and Teamwork. *In EAI Endorsed Transactions on Security and Safety*, 2013.

Kwang H Lee (2004). First Course on Fuzzy Theory and Applications. *Advances in Intelligent and Soft Computing*. Springer-Verlag Berlin and Heidelberg GmbH & Co. K.

Fatih Emre Boran, Serkan Genc, Mustafa Kurt, and Diyar Akay (2009). A Multi-criteria Intuitionistic Fuzzy Group Decision Making for Supplier Selection with TOPSIS Method. *Expert Systems with Appl*ications, Volume 36, Issue 8 (pp. 11363-11368). Elsevier.

F. Herrera, E. Herrera-Viedma, and J. L. Verdegay (1996). A Model of Consensus in Group Decision Making Under Linguistic Assessments. *Fuzzy Sets Syst* (pp. 73-87). Elsevier North-Holland, Inc.

Cengiz Kahraman, Da Ruan, and Ibrahim Dogan (2003). Fuzzy Group Decision-making for Facility Location Selection. *Inf. Sci. Inf. Comput. Sci* (pp. 135-153). Elsevier Science Inc.

Prashanth Rajivan (2013). CyberCog: A Synthetic Task Environment for Measuring Cyber Situation. Master's thesis, Arizona State University, Tempe, AZ, USA, 2011.

Developing behavioural markers for teams: a drilling team case study

Margaret T Crichton[a], and Scott Moffat[b]
[a]*People Factor Consultants Ltd*
[b]*People Factor Consultants Ltd*

ABSTRACT

Behavioural markers are commonly used to assess and provide performance feedback based on objective observations of behaviours. Most existing behavioural markers relate to behaviours exhibited by individuals working in a team environment. This paper describes team behavioural markers developed to capture team interactions by key roles in a drilling team during simulator-based well control exercises. Four key dimensions with example behaviours were identified for critical drilling team roles based on observations of 25 simulator-based exercises and subsequently trialled and used on another 160 exercises. These dimensions are Team situation awareness, Team decision making, Teamwork & communication, and Team workload & stress management. The behavioural markers are then used to provide feedback during debriefs of team performance. Although primarily developed for drilling teams, it is anticipated that this approach, and the resultant team behavioural markers can be modified for teams operating in other high hazard domains.

KEYWORDS

Situation Awareness/Situation Assessment; Education and training; team performance, observation; behavioural markers

INTRODUCTION

Simulator-based exercises form an integral part of training in many high hazard industries, especially aviation, medicine, nuclear power production, and maritime. These industries have been at the forefront in recognising the importance of human factors, in particular non-technical skills, to improve safety. The oil and gas industry is now also recognising the critical impact of non-technical skills, and the technological development of drilling simulators in oil and gas has led to increased opportunities to practise and test out both technical and non-technical skills in training courses. The effectiveness of training, especially simulator-based training exercises, relies hugely on the quality of the debrief provided following the exercise. Such feedback benefits by addressing both technical and non-technical knowledge and skills demonstrated during the exercise. The debrief itself then depends on the use of suitable metrics to assess a team's performance by providing objective feedback.

Non-technical skills taxonomies developed in aviation (NOTECHS: Flin & Martin, 1998), medicine (ANTS: Fletcher, Flin, McGeorge, Glavin, Maran, & Patey, 2003; NOTSS: Yule, Flin, et al, 2006), and maritime (Leadership behavioural markers: Devitt & Holford, 2010), have been designed to observe individuals working in a team environment. In each of these settings, a team is created to complete specific activities, such as flying a plane, carrying out medical operations, or sailing a ship. This team may be an ad hoc team, formed by bringing together a group of suitably qualified individuals to achieve the goal, or may be a long-standing existing team comprising individuals who work together on a routine basis, such as members of a shift. Although the members of the team are interdependent and goal focused, it is typically the performance of the individual in the team setting that is observed and debriefed.

Some behavioural markers have been developed to observe the performance of complete teams, primarily in medicine, such as Crisis management behaviours (Gaba, Howard, et al, 1998), and Observational teamwork assessment for surgery (OTAS: Healey, Undre, Sevdalis, Koutantji, & Vincent, 2006). Behavioural markers of surgical excellence have also been used for observations at the individual, team, and organisational levels (Carthey, de Leval, & Reason, 2000). Unlike many medical teams, a drilling team tends to be a semi-established team, but team members may change out due to holidays or sickness, and new members may join the team for short periods to provide specific expertise. Being a team of experts, however, does not imply that the team can be considered to be an expert team (Salas, Rosen, et al, 2007). Thus, a drilling team fulfils many of the characteristics of an action team in that expertise, information, and tasks are distributed across specialised individuals (Kozlowski, Gully, McHugh, Salas, & Cannon-Bowers, 1996).

This paper reports the development of a team behavioural marker system designed to be used for the observation of drilling team members during simulator-based exercises. Following a number of drilling rig well control events over the past 10 years, including the Macondo tragedy in April 2010 (Chief Counsel, 2011), the oil and gas industry, particularly the International Association of Oil and Gas Producers (IOGP), have recommended training in non-technical skills, such as a Crew Resource Management form of training course, for well operations team members (OGP 501, 2014). Such training typically addresses the performance of the individual in the team, yet while managing a well control incident, a drilling team relies on effective teamwork. In order to provide feedback about well operations team performance during simulator-based exercises, a specific team behavioural marker system was therefore required.

METHOD

Observations were carried out on 25 exercises conducted on a full-scale high fidelity drilling simulator which formed part of a well control 4-day training course. Each training course included five simulator-based exercises, therefore each exercise was observed five times over a 3-month period. The exercises involved interactions between a number of drilling team roles, including: Driller, Assistant Driller, Mud Logger, Drilling Supervisor, and Toolpusher. Other incidental roles, such as Offshore Installation Manager, Mud Engineer, and Rig Manager, were also included dependent on the number of team members participating in the training course. Team members primarily took on their own actual role during the exercise, the exception being if the Driller and Assistant Driller swapped out to provide the less experienced team member with the opportunity to practise on the simulator as part of the training course.

The exercise scenarios were created around specific challenging events that could arise requiring well control operations and evolved over real time in the simulator. The duration of each exercise was between 30 minutes and 3hrs 30 minutes. No interruptions were made to the flow of the exercise, unless a high risk situation appeared to be emerging and the team were struggling to cope. No such events occurred during the five exercises under consideration.

Following the procedure for observation-based non-technical skills identification described in Flin, O'Connor and Crichton (2008), a total of 70 specific observable interactions between team members were noted over the 25 exercises. Where multiple examples of the same behaviours were reported, only one interaction was included in the analysis. These interactions were then reviewed and categorised into team non-technical skills in the manner cited by Klampfer et al (2001) in that skills taxonomies should assess observable, non-technical behaviours that contribute to superior or substandard performance within a work environment. Two raters independently categorised the interactions and, after discussion, agreed the final four categories that form the team behavioural marker system.

RESULTS

Four categories of a team behavioural marker system were identified from the observations, namely: Team situation awareness, Team decision making, Teamwork & communication, and Team workload & stress management. Details of the Teamwork and communication category with relevant elements are shown in Table 1. Notably no leadership category was defined, as leadership was considered to be related to individual behaviours. Moreover a drilling team comprises a number of roles designated as leaders, such as the Toolpusher, Drilling Supervisor, and Driller, each of whom has specific responsibilities, and the interactions of these leaders as members of the team that affects overall team performance.

Category	Elements	Exceeds expectations	Meets expectations	Marginally below expectations	Well below expectations
Teamwork and communication *(All team members know and understand the contribution of their own role and that of others to achieve the team objectives. Exchange and confirm information in a timely and concise manner)*	Clarify roles and responsibilities Agree allocation and prioritisation of tasks Identify potential conflicts in team Show respect for individual contributions Participate, engage and support each other during activities Seek confirmation	Use open questions appropriately during and at the end of briefings Define and confirm roles and responsibilities for the current operation	Hold structured briefings Use open questions to check understanding of tasks prior to commencing Confirm and verify data being exchanged Use assertive style when asking questions or making challenges Clarify and agree	Hold unstructured or disjointed briefings (e.g. timeouts, tool box talk, handovers) Restrict participation or involvement of team members Shout information with no identified listener Use closed questions inappropriately	Use inappropriate questioning style Do not confirm data being exchanged Use inappropriate statements to respond (e.g. 'copy', 'roger') Carry out actions without discussing with others Do not speak up or speak out

	and check understanding Provide information clearly and concisely Communicate assertively		roles and responsibilities Seek and listen to specialist input Include relevant people in discussions Coach less experienced people during tasks	Confirm or verify data irregularly or infrequently Make challenges but not followed through with further debate	Use non-specific terminology (e.g. 'Slight', 'bit higher') Do not allocate or agree roles or responsibilities Carry out inadequate timeouts Call for briefings/ tool box talks/ timeouts but do not hold one Allow discussions to be interrupted inappropriately Do not avoid conflicts between team members occurring

Table 1. Example of team behavioural markers for teamwork and communication

Next, examples of the observed behaviours were rated on a four-point performance scale of: Exceeds expectations, Meets expectations, Marginally below expectations, and Well below expectations. Ratings were defined for the performance of the team as a whole. An observation sheet was developed listing the categories and relevant elements. The examples of behaviours were printed on the reverse of each observation sheet to enable observers to become familiar with them. It was not intended that these examples are used as a checklist as they should only be used to provide some guidance for team interactions.

The team behavioural markers have subsequently been piloted on a further 60 exercises and, through an iterative process, minor modifications have made to the system as appropriate. Over the next 8 months, the team behavioural marker system was used on a further 100 exercises. The system also formed the basis of reports submitted to Team Managers indicating the team's non-technical skills performance during the 4-day training course. The basis of the report was a graphical illustration of the team's performance using the four-point rating scale where Exceeds expectations was coloured emerald green; Met expectations was mid-green; Marginally below expectations was yellow; and Well below expectations was red. The team's performance during each exercise was thus easily identifiable (see Figure 1). The report being presented in this way has been well received by Team Managers, who can then use the report as a basis for checking for performance in the workplace.

Start of simulation → End of simulation

Team work & communication	Some learners interrupted others	Less experienced began to step forward and give their opinion	Excellent verification, mainly between driller and AD	Information not acknowledged and not repeated	Use of leading questions: "Did you see 400?"

Figure 1. Example of rated team behavioural markers for teamwork and communication

DISCUSSION

This paper set out to describe the development of team behavioural markers to act as a useful tool when providing feedback to teams following observations of exercises on a drilling simulator. During the well control training course, feedback was given at the termination of each exercise. Feedback lasted between 40 to 60 minutes, and covered both technical and non-technical skills performance. For many participants in the training course, this was the first time that they had actually received any targeted non-technical skills feedback; however, participants were generally extremely receptive to the information being provided to them.

It was notable that this feedback, during the first two or three exercises, predominantly highlighted areas for improvement, particularly in terms of the Teamwork & communication category. As shown in Table 1 above, the number of behaviours listed as Well below expectations for this category outnumbered all other categories and elements. Through practise and learning from the guided feedback, team performance over the last two exercises improved considerably. Team members were also observed coaching and prompting expected behaviours within the team during the later exercises.

Observers of team interactions have, to date, been experienced human factors professionals. In the future, however, a training programme can be designed to train those who might use the tool both during simulator-based exercises, or ideally, in the workplace. As Thorogood and Crichton (2014) comment, as well as providing training in non-technical (or Crew Resource Management) skills, the aim for the oil and gas industry should be to enhance existing workplace practices so that effective non-technical skills can be observed, coached and ultimately assessed in the workplace.

CONCLUSION

High fidelity simulators are frequently used in high hazard industries to train individuals and teams although this is generally focused on technical performance. Developing and using a team behavioural marker system offers an additional opportunity to enhance safety and performance by encouraging participants to review the team's behaviours and their own contribution to effective operations.

ACKNOWLEDGMENTS

The authors would like to thank the trainers and simulator exercise directors as well as the drilling team members who supported the development of this tool.

REFERENCES

Carthey, J., de Leval, M.R., & Reason, J.T. (2000) Understanding excellence in complex, dynamic medical domains. In *Proceedings of the International Ergonomics Association and Human Factors Society Triennial Conference*, 4. The Human Factors and Ergonomics Society Press, Santa Monica, CA

Chief Counsel's Report, (2011): *Macondo: the Gulf oil disaster*, National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling.

Devitt, K., & Holford, S. (2010). Development of resource management and leadership behavioural markers for the merchant navy. In *Proceedings of the Maritime Human Resource Solutions Conference*, 28-30 September, St John's Newfoundland, Canada.

Fletcher, G., Flin, R., McGeorge, P., Glavin, R., Maran, N., & Patey, R. (2003). Anaesthetists' non-technical skills (ANTS): Evaluation of a behavioural marker system. *British Journal of Anaesthesia*, 90, 580-588.

Flin, R., & Martin, L. (1998). *Behavioural markers for Crew Resource Management* (CAA Paper 98005). London: Civil Aviation Authority.

Flin, R., O'Connor, P., & Crichton, M. (2008) *Safety at the sharp end. A guide to non-technical skills.* Aldershot: Ashgate

Gaba, D. M., Howard, S., Flanagan, B., Smith, B., Fish, K., & Botney, R. (1998). Assessment of clinical performance during simulated crises using both technical and behavioural ratings. *Anesthesiology*, 89(1), 8-18.

Healey, A.N., Undre, S., Sevdalis, N., Koutantji, M., Vincent, C.A. (2006) The complexity of measuring interprofessional teamwork in the operating theatre. *Journal of Interprofessional Care*, 20, 485-495.

Klampfer, B., Flin, R., Helmreich, R.L., Hausler, R., Sexton, B, Fletcher, G., Field, P., Staender, S., Lauche, K., Dieckmann, Pl, & Amacher, A. (2001) *GIHRE (Group Interaction in High Risk Environments). Enhancing performance in high risk environments. Recommendations for the use of behavioural markers.* Behavioural Markers Workshop, Zurich.

Kozlowski, S. W. J., Gully, S. M., McHugh, P. P., Salas, E., & Cannon-Bowers, J. (1996). A dynamic theory of leadership and team effectiveness: Developmental and task contingent leader roles. *Research in Personnel and Human Resources Management*, 14, 253-305

OGP (2014). *Crew Resource Management for Well Operations teams*, Report No 501. International Association of Oil and Gas Producers

Salas, E., Rosen, M.A., Burke, C.S., Nicholson, D., and Howse, W.R. (2007). Markers for enhancing team cognition in complex environments: The power of team performance diagnosis. *Aviation, Space, and Environmental Medicine*, 78, 77-85

Thorogood, J.L. & Crichton, M. (2014) Threat and error management: The connection between process safety and practical action at the worksite. In *Proceedings of the SPE/IADC conference*, Fort Worth, TX

Yule, S., Flin, R., Paterson-Brown, S., Maran, N., & Rowley, D. (2006). Development of a rating system for surgeons' non-technical skills. *Medical Education*, 40, 1098-1104.

FROM LAB TO CAGE : TURNING OCCLUSION RESEARCH INTO SPORTS TRAINING

PETER FADDE
Southern Illinois University

fadde@siu.edu

ABSTRACT
Over the past 30 years, sport science researchers have used the method of *temporal occlusion* to investigate the perceptual-cognitive skills that allow athletes to defy limits of human perception when they return serves, block shots on goal, or hit pitched baseballs. However, the occlusion method has yet to be systematically used to train high-performance athletes. This study describes a 6-month program that used occlusion methods to train the perceptual-cognitive skill of pitch recognition in college baseball batters. The pitch recognition training program combined occlusion training using interactive computer software with live batting cage drills that also incorporated occlusion principles. The cooperating team's batting performance improved significantly, demonstrating that occlusion methods can be used to effectively train advanced perceptual-cognitive skills and thereby improve performance in sports. The combining of computer and *in situ* occlusion tasks has implications for training the recognition component of high-speed decision-making in sports and other domains.

KEYWORDS
Practical Application ; Judgment and Decision Making ; Expertise ; Learning and Training ; Education and Training ; Sports.

INTRODUCTION
Recognition-Primed Decision-Making (Klein, 1998) provides a useful model for understanding, and improving, ballistic sports skills such as returning a 130 mile-per-hour serve, blocking a penalty shot, and hitting a wicked googly or a nasty slider (Fadde, 2009). Such actions require athletes to select and execute a complex psychomotor response in time frames that challenge simple human reaction time. Sport science research has shown that these skills are not based upon super-human hand-eye coordination, reaction time, or vision but rather skill-specific schema built through massed experience (Williams & Ward, 2003). As David Epstein notes in *The Sports Gene* (2013), expert performers enjoy a software advantage rather than a hardware advantage. Software, in this case, consists of *perceptual-cogntive* skills that enable expert performers to rapidly recognize patterns and predict outcomes, thereby priming their impossibly fast reactions. The natural questions, then, are if and how expert perceptual-cognitive skills can be systematically trained in order to accelerate expertise.

This paper first describes the laboratory research method of *temporal occlusion* that was developed by sport science researchers as away to isolate and measure perceptual-cognitive skills. The paper then describes an extended sports training program in which computer-based occlusion activities were mixed with *in situ* occlusion activities and implemented with a high-level sports team. The study addresses two critical issues that have previously limited the application of experimentally validated occlusion methods for training perceptual-cognitive skills: 1) transfer from training to performance, and 2) implementation in authentic settings with advanced performers.

Occlusion Methods to Study Perceptual-Cognitive Expertise
Sport scientists have developed a variety of occlusion tasks in which subjects view an opponent's action and categorize the action (e.g., type of tennis serve) or predict the outcome. The view is masked (spatial occlusion) or cut off (temporal occlusion) in different ways to remove perceptual information. If removing a particular piece of perceptual information results in a notable decrement in expert subjects' performance advantage over novices then the occluded information is deemed to have been important to the experts' perceptual advantage

(Williams & Ward, 2003). For instance, when masking a particular part of the bowler's body leads to a reduction in expert cricket batters' ability to "guess" the type of ball being delivered then researchers assume that some of the experts' perceptual advantage is gained by attending to that part of the body during the bowler's run-up motion (Müller & Abernethy, 2012).

In the most commonly used type of occlusion in sport science laboratories, the visual display of an opponent's action is cut off at various points of time during the action. In a typical temporal occlusion study of tennis return-of-serve video clips of a server were variously cut off before the ball was struck, at the moment of racquet-ball contact, and very shortly after contact. Study participants with different degrees of tennis expertise were tasked with categorizing the type of serve while viewing occluded video clips. Expert tennis players were better able to categorize serve type based on less visual information (Scott, Scott, & Howe, 1998). Sport science researchers have used both the findings of occlusion research and the occlusion method itself to train perceptual-cognitive skills. For instance, Farrow, Chives, Hardingham, and Sauces (1998) demonstrated the effectiveness of video-based occlusion training on the return-of-serve skill of intermediate tennis players.

Occlusion Testing and Training of Baseball Pitch Recognition

Temporal occlusion has been used to both test and train *pitch recognition* as the perceptual-cognitive component of baseball batting. Figure 1 shows a typical laboratory-based temporal occlusion task used by researchers to confirm experts' perceptual-cognitive advantage (e.g., Paull & Glencross, 1997) and also to train the same perceptual-cognitive skills (Burroughs, 1984; Fadde, 2006). Figure 2 shows a computer-based version that is also be used for both research and training purposes.

Figure 1. Video-Simulation / Occlusion in lab

Whether in a laboratory setting or on a laptop computer, occlusion training is usually presented in the form of *video-simulation* in which users respond to a video display by inputting a choice (e.g., Pitch Type) or prediction (e.g., Pitch Location) via keyboard, mouse, touch, or voice. Typically, however, users are not required to perform a psychomotor skill such as returning a serve or hitting a pitch. This de-coupling of the perception-action link allows researchers to isolate the perceptual-cognitive component of performance for testing or training purposes but also raises questions of ecological fidelity (Bootsma & Hardy, 1997). The part-task approach of video-simulation contrasts with whole-task video-based simulators, such as depicted in Figure 3, in which a pitching machine propels a ball through a video projection screen to simulate batting.

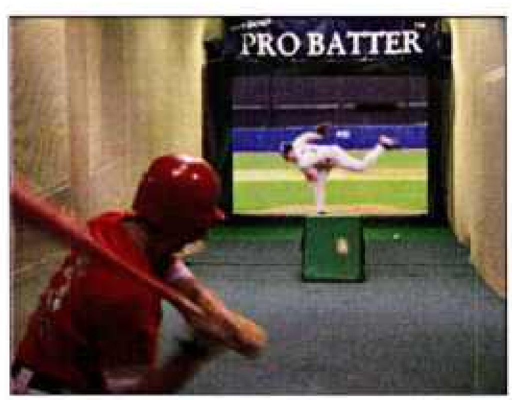


Figure 2. Video-Simulation (courtesy AxonSports) Figure 3. Video simulator (courtesy ProBatter)

The video simulator shown in Figure 3 is capable of "throwing" a variety of pitches such as fastball, curveball, and changeup. However, the video of the pitcher does not change in relation to the type of pitch delivered thereby denying the user authentic pitch release cues. In essence, the simulator has higher fidelity for the whole task of baseball batting but video-simulation has higher fidelity for the partial task of pitch recognition.

As a part-task recognition-only training method, video-occlusion (temporal occlusion in a video-simulation format) offers a high degree of instructional efficiency (Fadde, 2009). Baseball batters can train an important component skill on a portable device during travel or rehabilitation from injury. However, the issue of transfer of part-task perceptual-cognitive learning to whole-task psychomotor performance looms large.

Occlusion Training and Transfer to Performance

Transfer of learning comes in many forms and terminology is not always consistent. A useful delineation can be made between near transfer and far transfer. *Near transfer* refers to trainees' performance in a video-occlusion task compared to their performance in an *in situ* version of the same task. For example, Burroughs (1984) used his patented Visual Interruption System (Figure 4) to re-create a "live" version of the video-occlusion pitch recognition task that he used to train college baseball players.

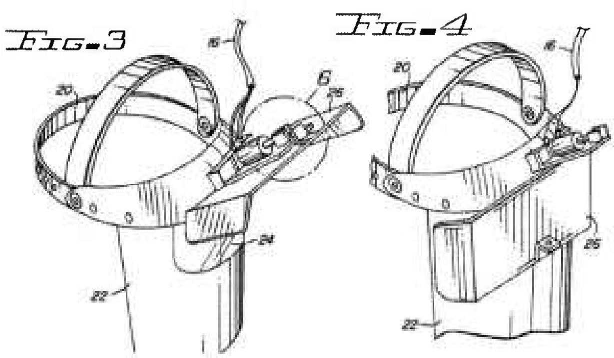


Figure 4. Visual Interruption System (patent illustration)

Figure 5. Occlusion Spectacles (courtesy Translucent Technologies)

The baseball players participating in Burroughs' study first received video-occlusion training in a laboratory setting. They then moved to a baseball field where a pitcher threw full-speed pitches and VIS was used to occlude the batters' vision after a short amount of ball flight. VIS used the landing of the pitcher's front foot on a force pad to send an electronic signal to a hinged visor that would snap down, blocking the batter's vision. Batters identified the type of occluded pitch or predicted the location of the pitch, just as they had when watching occluded video clips of pitches in the laboratory. The study demonstrated near transfer of learning gains made in a video-occlusion task to an analogous *in situ* occlusion task. Occlusion spectacles (Figure 5) have been used in a similar way by researchers conducting video-occlusion testing and training in cricket (Müller & Abernethy, 2012).

Having participants or trainees perform a simplified version of the full psychomotor performance task represents a different type of transfer. For example, Scott, Scott, and Howe (1998) used video-occlusion to train tennis players to recognize types of serves. They then had the players perform a return-of-serve task on court with a "live" server. Players scored increasing points for making contact with the serve, returning the serve over the net, or returning the serve into the server's court. The researchers' assumption was that a higher score indicated that trainees were successfully applying the skill of picking up early cues that they had practiced during video-occlusion training. Researchers have sometimes combined *in situ* occlusion tasks (as Burroughs) and on-court/field representative tasks (as Scott, Scott, & Howe) approaches by, for example, having cricket batters not only identify but also attempt to strike bowled balls while having their vision cut off with occlusion spectacles (Müller & Abernethy, 2006).

Far Transfer of Occlusion Training

The transfer of perceptual-cognitive training gains to psychomotor performance of the full skill in match situations can be thought of as far transfer. As in all areas of training, transfer to performance can be very challenging to measure. However, a benefit of investigating baseball batting is that performance is systematically measured by established statistics. Fadde (2006) used laboratory-based video-occlusion methods (see Figure 1) in a training program intended to improve the pitch recognition skill of college baseball batters. Batters on the same college baseball team were randomly assigned to occlusion training and control groups.

During the team's winter practice sessions at an indoor facility, players in the training group left the practice field to complete individual 15-minute video-occlusion training sessions. Upon completing video-occlusion sessions, players returned to team practice that often included situational batting against full-speed pitching. The actual treatment, therefore, was a combination of video-occlusion training and "live" batting in the context of organized team practice.

The effectiveness of the pitch recognition training program was determined by comparing the batting statistics of players in the training and control groups during the team's 18-game pre-conference schedule. Batters in the training group had higher Batting Average, On-Base Percentage, and Slugging Percentage – the batting statistics generally considered to represent batting skill (Weinberg, 2014). Rank correlation of batters was used to determine statistical significance of the differences between training and control groups, which was statistically significant on the measure of batting average ($p<.05$).

Limitations of Occlusion Training Studies

Training-based research studies are usually experimental in design, attempting to isolate and validate the effectiveness of training methods. These studies purposefully limit the duration and context of experimental training interventions in order to strengthen experimental control. The internal validity that is maximized by controlled experimental designs, however, limits the external validity of the studies. Wider adoption of occlusion methods for training high-performance athletes depends upon the implementation and study of perceptual-cognitive training programs in authentic settings with high-level performers.

OCCLUSION TRAINING OF PITCH RECOGNITION IN A NATURAL SETTING

The study described here implemented and evaluated a training program that used occlusion principles to train the perceptual-cognitive skill of pitch recognition. The ability of expert batters to pick up early cues in the pitcher's delivery and early ball flight is not only well established by sports science research as a differentiating skill of expert batters (Paull & Glencross, 1997) but is also recognized as a valuable skill by many college and professional baseball teams. However, it has not generally been considered to be "coachable" (White, 2014, June 4). This study investigates whether occlusion training methods, when incorporated into the routine practice activities of a NCAA Division I college baseball team, would transfer to improved batting performance.

Integrating Computer and Batting Cage Pitch Recognition Drills

The pitch recognition training program involved players individually using a computer application (see Figure 2) created by *AxonSports* that presented temporal occlusion drills in a format that combined drill-and-practice methodology with dynamic testing in which the occlusion point of video pitches was automatically shortened as players achieved target scores. Players could choose to work on Pitch Type or Pitch Location drills and could choose among three video pitchers.

The pitch recognition training program also included several "live" batting cage drills that added a layer of pitch recognition to traditional batting drills. For instance, rather than simply hitting the ball off a tee, batters would watch a teammate or coach deliver a mock pitch from behind a protective screen and hit the ball off the tee only when they recognized the mock pitch as a designated type of pitch (e.g., fastball, curveball, changeup). As shown in Figure 6, the net occluded the pitch very much as the computer program did by editing to black.

Figure 6. Net Occlusion Drill

Another "live" drill was similar to the *in situ* near-transfer tasks used in occlusion training research (e.g., Burroughs, 1984). Instead of measuring transfer, however, this drill was intended to facilitate transfer by replicating the computer occlusion drills with live pitchers. While the team's pitchers were practicing pitching in the bullpen (a designated area at baseball fields where pitchers warm up), the batters would *stand in*. That is, a batter would take a normal position in the batter's box but would not swing his bat. In traditional stand-in drills, batters are tasked with tracking the pitch into the catcher's mitt.

Figure 7. Bullpen Stand-In Pitch Recognition Drill

The *Stand-In Pitch Recognition* drill interjected occlusion into this routine drill by instructing batters to call the type of pitch out loud before the ball hit the catcher's mitt. Batters also predicted whether or not a pitch would be in the strike zone, calling out "Yes" to indicate the pitch would be a strike and "No" to predict the pitch would not be a strike. With a typical pitch reaching the catcher in less than 500 milliseconds, the requirement that batters verbalize their pitch call enforced early recognition – a variation termed *attention occlusion.*

Results of Pitch Recognition Training Program

The effectiveness of the Pitch Recognition Training Program was measured in terms of batting performance in conference games. Official NCAA batting statistics were used. In this case study all of the batters on the team were trained and the team's mean batting statistics were compared between the 2013 and 2014 seasons. As summarized in Table 1, the trained team showed consistent and substantial improvement from the 2013 season to the 2014 season. While it is not possible to attribute the team's batting performance gains to the pitch recognition training program, both the coaches and the researcher wanted to know whether the observed improvements were "beyond the reasonable expectation of a good team getting better." To address this question the participating team's batting statistics were compared with those of a comparable conference team that had similar batting statistics in the 2013 season but did not receive pitch recognition training. Like the trained team, the nop-training team returned 8 out of 9 batters in its starting lineup from 2013 for the 2014 season.

Table 1 displays the teams' batting statistics in 2013 to 2014. Batting average, on-base percentage, and slugging percentage are considered to be a basic profile of batting performance (Weinberg, 2014). Base-on-balls (BB), strikeouts (K), and BB/K ratio are considered to represent "good eye" or plate discipline (Panas, 2010). Runs-per-Game is highlighted in Table 1 as the most basic measure of team batting performance. Base-on-Balls/Strikeouts (BB/K) ratio is highlighted as the statistic most closely associated with plate discipline. For contextual purposes (not specific to this study), general benchmarks of excellence for these statistics include: Batting Average (.300), On-base Percentage (.380), Slugging Percentage (.450), and BB/K (.50). The measure of Strikeouts is reverse scored because fewer strikeouts represent better batting performance.

As can be seen in Table 1, the no training team showed modest improvement in most team batting statistics from 2013 to 2014, as would be expected from a team returning most of the starting players from the previous year. However, the training teams' batting statistics showed consistent and often substantial improvements from 2013 to 2014, well beyond expectations for a good team returning most of its starting lineup.

Table 1. NCAA Batting Performance Statistics for Training and No Training Teams

	Training Team 2013 2014 Change			**No Training Team 2013 2014 Change**		
Runs Per Game		8.6	**48%**		6.8	
Batting Average		.326	14%		.304	
On-base Percentage		.407	9%		.383	
Slugging Percentage		.468	20%		.464	
Home Runs		25	127%		27	
Base-on-Balls		140	30%		124	
Strikeouts (K)		182	16%		200	
BB/K Ratio		.77	**54%**		.62	

To test whether the differences between the two teams went beyond face-value and were statistically significant, I compared the changes in both teams' ranking among conference teams for 2013 and 2014 batting statistics. As in Fadde's 2006 study, statistical significance was determined by comparing the ranking of the training and no training teams' within the conference and applying the Mann-Whitney *u*-test of rank correlation, scaled for small *n*. Figure 8 displays Rank data. With eleven teams competing in the conference, the top rank-based score is designated as "11" and the bottom rank is designated as "1" on the graph.

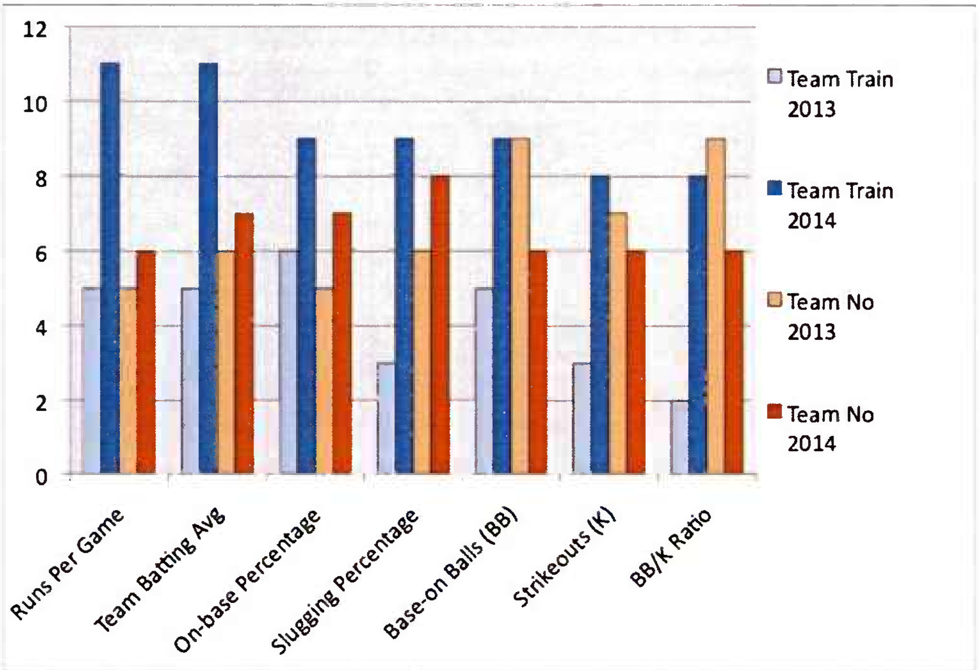


Figure 8: Ranking of Training and No Training Teams' Batting in Conference (11=best; 1=worst)

Applying a one-tailed analysis with alpha of $p<.01$, the season-to-season change for pooled rankings in the batting statistics of the team receiving pitch recognition training is significant ($p=.0005$) while the change in pooled batting statistics of the comparison team is non-significant ($p=.4364$).

CONCLUSION

Occlusion methods originally developed by sports science researchers to verify and locate the sources of expert advantage in perceptual-cognitive skills have been shown, through time-limited experimental implmentations, to be effective training methods. However, occlusion methods had not previously been systematically applied to the training of high-performance athletes. This case study addressed this need by implementing an occlusion-based pitch recognition training program that targeted already high performing athletes. The training program used temporal occlusion embodied in a video-simulation on a laptop computer and also incorporated occlusion principles into "live" drills that essentially simulated the computer simulation.

The attention occlusion method of calling the pitch out loud before it hits the catcher's mitt, which was developed for the training program in this study, is now being used for training pitch recognition in at least one major league baseball organization (White, 2014, June 4), and the study has implications for targeted, part-task, occlusion-based training of the recognition component of high-speed decision-making (Fadde, 2009) in sports and a variety of other performance domains.

REFERENCES

Bootsma, R. & Hardy, L. (1997). Perception and action in sport: Half-time comments on the match. *Journal of Sports Sciences*, *15*, 641-642.

Burroughs, W. (1984). Visual simulation training of baseball batters. *International Journal of Sport Psychology, 15*, 117-126.

Epstein, D. (2013). *The sports gene: Inside the science of extraordinary athletic performance*. New York: Penguin.

Fadde, P. J. (2006). Interactive video training of perceptual decision-making in the sport of baseball. *Technology, Instruction, Cognition, and Learning, 4*(3), 265-285.

Fadde, P. J. (2009). Instructional design for advanced learners: Training expert recognition skills to hasten expertise. *Educational Technology Research and Development, 57*(3), 359-376. doi: 10.1007/s11423-007-9046-5

Farrow, D., Chives, P., Hardingham C. & Sauces, S. (1998). The effect of video-based perceptual training on the tennis return of serve. *International Journal of Sport Psychology, 29*, 231-242.

Klein, G. (1998). *Sources of power : How people make decisions*. Cambridge, MA : MIT Press.

Müller, S. & Abernethy, B. (2006). Batting with occluded vision: An in-situ examination of the information pick-up and interceptive skills of high and low skilled cricket batsmen. *Journal of Science and Medicine in Sport, 9*, 446-458. doi:10.1016/j.jsams.2006.03.029

Müller, S. & Abernethey, B. (2012). Expert anticipatory skill in striking sports: A review and a model. *Research Quarterly for Exercise and Sport, 83*(2). 175-187.

Panas, L. (2010). *Beyond batting average*. Retrieved from http://www.lulu.com/us/en/shop/lee-panas/beyond-batting-average/paperback/product-10663690.html

Paull, G. & Glencross, D. (1997). Expert perception and decision making in baseball. *International Journal of Sport Psychology, 28*, 35-56.

Scott, D., Scott, L. & Howe, B. (1998). Training anticipation for intermediate tennis players. *Behavior Modification, 22*(3), 243-261.

Weinberg, N. (2014, May 26). How to evaluate a hitter, sabermetrically. *Beyond the Box Score*. Retrieved from http://www.beyondtheboxscore.com/2014/5/26/5743956/sabermetrics-stats-offense-learn-sabermetrics

White, P. (2014, June 4). Hard knock life: Astros turn to plate discipline for success. *USA Today*. Retrieved from http://www.usatoday.com/story/sports/mlb/2014/06/03/houston-astros-build-plate-discipline-from-the-ground-up/9934613/

Williams, A.M. & Ward, P. (2003). Perceptual expertise: Development in sport. In J. L. Starkes & K. A. Ericsson (Eds.), *Expert performance in sports: Advances in research in sport expertise* (pp. 219-247). Champaign, IL: Human Kinetics.

Distributed Sensemaking: A Case Study of Military Analysis

Simon ATTFIELD[a], Bob FIELDS[a], Ashley WHEAT[a], Rob HUTTON[b], Jim NIXON[c] and Andrew LEGGETT[b]
[a]*Interaction Design Centre, Middlesex University, UK*
[b]*Advanced Technology Centre, BAE Systems, UK*
[c]*Safety and Accident Investigation Centre, Cranfield Univerisity, UK*

ABSTRACT
This paper develops a perspective on understanding the phenomenon of sensemaking by taking as the unit of analysis, not the mind of an individual sensemaker, but an assembly of people and artefacts, potentially distributed physically, socially and over time. The paper reports an observational study of military analysts that explores how a sensemaking task can be understood in terms of the distribution of task-relevant representations across internal, mental media, and external, physical ones. The design and interactional properties of such external representational medial has a profound effect on the properties of the combined distributed sensemaking system. A rich account of the Distributed Sensemaking work involved is presented that draws inspiration from existing models of sensemaking as well as the distributed cognition approach of Hutchins.

KEYWORDS
Sensemaking; externalized/embedded cognition; military; Distributed Cognition; Distributed Sensemaking.

INTRODUCTION

Sensemaking has been described as a process of comprehension (Klein et al., 2007), and of finding meaning from information (Weick, 1995). Representation is central to sensemaking. In sensemaking we build 'pictures' or representations of aspect of the world using data that we receive about it. A number of theoretical accounts of sensemaking have been proposed including the Data/Frame model (Klein et al., 2007), Weick's analysis of sensemaking in organisations (Weick, 1995), Pirolli and Card's analysis of sensemaking by intelligence analysts (Pirolli and Card, 2005), Russell et al's Learning Loop Complex model (Russell et al., 1993) and Dervin's Sense-making Methodology (1983). For some models, the focus is on processes that surround representations 'in the head', in the form of beliefs, or 'mental models'. But sensemaking frequently also involves the use of representations which are 'in the world'. These may take many forms including lists, maps, charts, pictures or reference information.

Representations in the world presumably change sensemaking in some way, depending on things like visual form and interaction properties. Testament to this expectation lies in the extensive research into representational methods and interactive tools for helping during sensemaking tasks. But of the available models and theories, few seem to engage with this in any depth. And they seldom explore how and why representational artefacts affect sensemaking processes and outcomes.

With this in mind we consider a sensemaking task which involves the use of external representational artefacts, analysing it in a way which attends to how these artefacts support, mediate and enhance cognition during selected parts of the task. We use as our example an observational study of military analysts involved in a training exercise. Our approach is influenced by the perspective of Distributed Cognition (DC). DC has its foundations in the cognitive ethnography of Hutchins (1995a). Hutchins argued that cognitive processes are best understood when we see them as distributed across socio-technical work-systems. DC argues that a complete explanatory account not possible without considering how it is distributed across materials, time and people. Our interest is to contribute to a better understanding of Sensemaking as Distributed Sensemaking, with a particular interest in how external representations support reasoning.

In the next section we give some background on sensemaking and on Distributed Cognition. In section 3, we present the case-study and in section 4 we then describe the case study design . In section 5 we report findings with a focus on how external artefacts affected cognition, and in section 6 we consider the implications of the findings

BACKGROUND

Sensemaking

Sensemaking concerns the ways in which we use information to construct interpretations of the world around us. Different theories of sensemaking have drawn attention to different aspect of this process and considered it in different contexts. Weick (1995) for example was concerned with the forces that act on sensemaking within organisational settings in which individual and social sensemaking were inextricably linked. Dervin (1983) has been concerned with how sensemaking relates to information seeking and information needs, Russell et al (1993) described how representational schema change to accommodate ill-fitting information and Pirolli and Card (2005) described a sensemaking process model based on a task analysis of intelligence analysts.

The data-frame model (Klein et al., 2007) is helpful from the perspective of offering a fairly detailed account of cognitive processes involved in sensemaking. It refers to two kinds of entity, data and frame, which interact dynamically during sensemaking. Data are aspects of the world that a sensemaker experiences as they interact with it. A frame is a representation, which stands as an account of that situation. For example, it might include a doctor's beliefs about a patient's medical condition, a pilot's understanding of his current location and heading, or a warship captain's beliefs about the objectives (and potential threat) of an approaching aircraft. In this sense, a frame acts as both interpretation and explanation of data.

The theory presents sensemaking as a continual process of framing and re-framing in the light of data. As we encounter a new situation a few key cues, or anchors invoke a plausible frame as an interpretation of that situation. Active exploration guided by the frame is then used to elaborate it or challenges it by revealing inconsistent data. By extending further than the observed data, a frame offers an economy on the data required for understanding, but also sets up expectations for further data that might be available. Hence a frame can 'direct' information search and in doing so reveal further data that changes the frame. An activated frame acts as an information filter, not only determining what information is subsequently sought, but also affecting what aspects of a situation will subsequently be noticed.

The particular frame that is activated may depend upon a number of things including: available cues, workload, motivation, and also the sensemaker's repertoire of frames. People have different frame repertoires based on prior experience with this underpinning a distinction between experts and novices. A frame creates expectations and violations can come as a surprise, bringing a frame into question and provoking re-assessment of the current 'understanding'. However, a frame can be maintained in the light of conflicting data, including in the case of confirmation bias and potentially unreliable data. From a representational perspective, we regard sensemaking as a process which demands coherence between different levels of representation of a given domain or area. Something 'makes sense' when what we see it is consistent with general beliefs we hold about that situation or situations like it.

Distributed Cognition

Hutchins argued the need for cognitive science to be broadened to include whole cognitive environments of which the individual is a part (Rogers, 2012). Compared with a more traditional view of cognition, distributed cognition extends the unit of analysis to a concern with the ways in which cognitive processes transcend boundaries of the individual, taking into account an interplay between people, internal and external representations, and the use os artefacts which are said to form part of a wider 'cognitive system' (Hutchins, 1995a; Rogers, 2012). It argues that an explanatory account of cognition which fails to include such factors is incomplete. Hutchins and colleagues (Hollan, 2000) propose that cognition can be distributed in a number of ways, describing the distributed cognition approach as three 'tenets'. These are: *Socially Distributed Cognition*: which describes cognitive tasks as being distributed across individuals acting together; *Embodied Cognition*: describing the distribution of cognitive tasks across internal and external resources and representations; *Culture and Cognition*: which describes cognitive processes as being shaped by cultural practices and ecologies.

Hutchins' approach has been applied in a number of cognitive systems in situated settings, including ship navigations (Hutchins, 1995a), airline cockpits (Hutchins 1995b, Hutchins and Klausen, 1996), air traffic control (Halverson, 1995) and emergency medical dispatch (Furniss and Blandford, 2006). A notable example is *How a Cockpit Remembers Its Speeds* (1995) in which Hutchins takes a socio-technical system—namely an airline cockpit—as the unit of analysis. He demonstrates how the entire cockpit of a commercial airliner performs cognitive tasks; computing and remembering airspeeds and wing configuration in preparation for an approach to landing. This was done by carrying out a number ethnographic observations of cockpit aircrews (pilot and co-pilot) flying mid-sized jets. An analysis of this shows how the memory of a cockpit is made up of not just individual pilot's memories, but much of the computation and processing required for flying a commercial airliner is carried out externally, where the pilots themselves are components of a larger cognitive system. Hutchins theorises about human's ability to design and manipulate our environments in order to complete cognitive tasks. One such example from the cockpit is the speed bug. Speed bugs are indicators that can be manually positioned on the cockpit's airspeed dial. Pilots position the speed bugs according to values illustrated on speed cards showing desired speeds at specific times in the approach to landing an aircraft according to certain conditions. In a descent the pilots use speed bugs as indicators of desired airspeeds at various points and as a means of cross-checking to ensure the aircraft is configured correctly.

A key conclusion made by studies in distributed cognition is an account of the interdependencies drawn between actors and artifacts in their working environments. A distributed cognition approach analysis is able to provide multi-level accounts of the elements making up a distributed cognitive system (Rogers, 2012). Interpreted through a "cognitive ethnographic" lens (Hutchins, 1995a, pg.371), studies are able to describe how abstract information structures (Wright, Fields and Harrison, 2000) are propagated and translated through various representational states, and the different media and resources that are used. Essentially, Cognitive Ethnography is a descriptive enterprise which aims for descriptions of the cognitive task world (Hutchins, pg.371). A Distributed Cognition account describes how people in naturalistic settings appropriate external cognitive resources, given their particular properties and affordances, in the service of strategies which implement useful computation. In the next section we explore this specifically in the context of sensemaking, taking as our example a study of military communications intelligence analysis.

CASE STUDY

We observed a group of ex-military analysts tackling an intelligence analysis training scenario. The scenario featured a simulated military landing on the South coast of England. Figure 1 shows how, within the scenario, the analyst's role fitted within a broader intelligence cell. Within the cell, Interceptors (left) are radio operators in the field who pick up radio broadcasts and distill information from these and send it to the Direction Finder (or 'Pilot'). The job of the Direction Finder is to use information from multiple Interceptors to triangulate locations of units in the field and to compile Tactical Tip Off (TTO) reports to send to the Analyst. The TTO includes the location information as well as the radio frequencies used, details of call signs and excerpts from the communications. The Analyst uses these reports to build a situation picture and provide periodic reports to a Supervisor. Priorities for monitoring and requests for information can be communicated upstream.

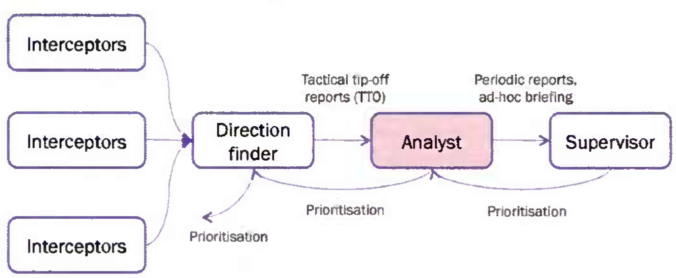


Figure 1. The analyst's role in the context of other roles within an intelligence 'cell'. For the purposes of this study we focussed on the analyst position.

The analysts worked using a computer with four displays in a two-by-two formation (see Figure 2). Software used by the analysts included the EW Training & Mission Support Tool (EWMST)[8] (top left screen) which enabled the mapping of assets in the field and the assignment of properties such as radio frequency or call signs. They also used IBM i2 Analyst's Notebook (top right screen) to create a network graph depicting command structure relationships. They used Microsoft Word to view TTO reports (bottom left screen) and create intelligence reports (bottom right screen) to send to the supervisor. They also used instant messaging software to communicate and exchange files with the Direction Finder and Supervisor.

The analysts were also given a set of printed materials known as 'working aids'. These included the 'Radio Equipment' table, detailing information about radio types in use; the 'Radio Procedures-Callsigns' table describing call sign procedures, equipment lists, maps, and the Order of Battle (ORBAT) for the adversary force. An ORBAT describes the known structure of an army including information about command structures and hierarchies, divisions, units and formations, personnel and equipment.

[8] EWMST is proprietary software developed by MASS consultants Ltd (UK).

Figure 2. The analyst's work station - shows the analyst (sitting) and the supervisor (standing). The training exercise was devised and run by MASS Consultants Ltd.

For the exercise, the analyst, pilot and supervisor were in the same room. Two runs of the scenario were observed with two different participants in the analyst position. A video camera was used to record the analyst's workstation from an over-shoulder perspective (as in Figure 2) and recordings were made of the analyst's four screens using screen capture software. A secondary camera recorded a wider view of the room. A log was created of the instant message communications between pilot and analyst and between supervisor and analyst. Audio recordings were made of their conversations and added to the video footage. The analysts were stopped by the researcher approximately every 15 minutes and asked to give an account of the current situation and their activities, which was also recorded. At the end of the first run, a de-brief interview was conducted with the analyst, and at the end of the second a debrief interview was conducted with the supervisor. This final debrief outlines the 'normative' process that analysts would ideally follow with the information and tools provided.

Audio, video and screen recordings of the exercise were transcribed as written narratives with the support of the data collected in instant message logs. This preserved continuity and avoided fragmentation of the data allowing us to clearly interpret the flow of information over time, and the translation of information through various resources and representations used by the analyst. We also treated the briefing from the supervisor in the same way, writing a narrative of the 'ideal' process analysts would take.

OBSERVATION FINDINGS

The analyst's job was to combine information provided in TTO reports with background information in tables and charts to draw inferences about the adversary force, to use this to generate a 'situation picture' and to keep the supervisor informed. Each TTO provided information about intercepted communications at a given frequency. Communications at a given frequency would correspond to one part of the opposing force—each part being a sub-network within a larger command structure. Each sub-network consisted of a command node and a number of subordinate nodes, although from the communications it wasn't necessarily obvious which was which.

When a TTO arrived (via the instant messaging tool), the analyst would usually begin by plotting the communicating entities mentioned in the TTO on the map (top left, Figure 2). They did this by invoking entity icons in the software and for each one inputting latitude and longitude information into a properties dialogue. With entities positioned visually on the map, the analyst would then use other information in the TTO to draw further inferences.

A first step was to attempt to ascertain the level of command of the communication, that is, to work out at what level within in the overall command structure the communicating entities were. The analysts were provided with a number of tables which could help them to do this. The first was a 'Radio Equipment' table; this linked known enemy radio types (14 types) to their operating frequency ranges, modes, level of command at which they were used and any associated remarks (see Figure 3). The analysts were able to use this table to draw inferences about level of command from the frequency of the communications.

In an early example in the scenario, a TTO reported communications on a frequency of 3.55MHz AM. Using the table, the analyst reviewed the operating frequency range of each radio type to see whether it included 3.55MHz. This was done by visually scanning each row in turn assessing whether or not the target frequency fell within the staged range (e.g. does 3.55MHz fall within "1.25 - 4.5 MHz"?). Cognitive complexity arises for the analyst when there are several possible matching rows in the table. The table was printed on paper and this afforded the approach, used by the analysts, of considering each row (and therefore radio) in turn, and using a pen to strike through rows where the frequency fell outside the range. Where the frequency fell within range, it was left unmarked. In the case of the 3.55MHz transmissions, this eliminated 12 or the 14 possible radio types and left two. The two possibilities were an P-404 which would imply Regiment-to-Battalion communication, and a P-434 which would imply Division-to-Regiment communication.

Radio Designation	Frequency Range	Mode(s)	Level of Command	Role	Remarks
P-404	1.25 - 4.5 MHz	FM / AM	REGT > BN	HF CNR	
P-407	28 - 50 MHz	FM / CW	BN > COY	VHF CNR	
P-411	20 - 48 MHz	FM	REGT > BN	VHF CNR	Command Vehicles, Listening Stations & Ground Launch / Radar stations
P-423	20 - 52 MHz	FM	REGT > BN	VHF CNR	Mainly used in Tanks and Armoured Vehicles
P-426	50 - 54 MHz	FM	BN > COY	VHF CNR	Manpack
P-429	1 - 18 MHz	AM / SSB / CW	BN > COY	HF CNR	
P-430	1.5 - 13.5 MHz	SSB	REGT > BN	HF CNR	
P-434	1 - 30 MHz	AM	DIV > REGT	HF CNR	
P-443	2 - 28 MHz	AM / SSB / CW	DIV > REGT	HF CNR	
P-445	52 - 62 MHz	FM	DIV > REGT	VHF CNR	
P-603	48 - 51 MHz	FM	Various	Radio Relay / Rebroadcast	Seen at all levels of command
P-707	52 - 58 MHz	FM	Various	Radio Relay / Rebroadcast	Seen at all levels of command
P-709	100 - 150 MHz	AM	ARMY > DIV > REGT	UHF G2A / CNR	Used for Ground to Air (G2A) comms and for Army - Div links

Figure 3. Radio Equipment table showing radio designations, frequency ranges, level of command and other information

To narrow the possibilities further, the analysts used the 'Radio Procedures-Callsign' (Figure 4) table that linked different forms of callsign (5 forms, e.g. 2-letter, 3 figures) to a range of factors including levels of command. In the example, the TTO indicated 6 call signs in use on the sub-network, each with two letters. Again, the analysts struck out rows that the call sign form excluded. In the example, they were able to eliminate all but one entry corresponding to communication at 'regiment and below'.

Type of Callsign	Example	Level/type of use
4-Symbols	S21A	Army > Division
2-Letters	NE	Regiment & below
3-Figures	245	Regiment & below
Word + 2-Figures	LION-17	Battalion > Coy (or equivalent)
Word	TORPEDO	Division > Regiment (or equivalent)

Figure 4: Radio Procedures-Callsign table connecting callsign format to organisational usage

With both of these conclusions in hand, the analysts could combine results by effectively performing a Boolean conjunction. The first table gave '(regiment-to-battalion) OR (division-to-regiment)' and the second gave 'regiment and below'. Given a hierarchy of Army > Division > Regiment > Battalion > Company, the Boolean operation (performed mentally): ((regiment-to-battalion) OR (division-to-regiment)) AND (regiment and below)' meant that the result could only be 'regiment-to-battalion'.

Having established level of command the analysts could use this information as a foothold, rather like a climbers piton, to find out other information. They reviewed the TTO for content that might help to determine the type of regiment. One of the communications used an apparent codeword as if referring to a particular piece of equipment (i.e. 'x are ready'). Finding the codeword in an Equipment table showed that it was an artillery piece. From this they could infer that the communication was between regiment to battalion within an artillery regiment. This interpretation was reinforced when they considered other message extracts including terms like 'FP', which could mean 'firing point' and FO which could mean 'forward observer'.

Using the regiment type, another piton, the analysts were able to then associate call signs with specific military units. This was done using an ORBAT which shows the elements within an army in their hierarchical organization. Finding an artillery regiment which has a battalion with the artillery piece, the analysts were able to associate it with the call sign which discussed it. They were then able to make informed guesses and a process of elimination to determine which battalions corresponded with the call signs.

DISCUSSION

The case study represents sensemaking in action. Information about the world, or rather, a part of the world, comes to the analyst providing cues for how he might begin to develop a frame or 'situation picture'. Combined with background knowledge, the cues support the analyst in making inferences about the information to develop a situation picture, or 'frame', which may be useful, given a set of interests and goals. Also, cognition is mediated by external representational artefacts and it would be difficult to explain the outcomes without reference to them, and so it is an example of distributed cognition.

In the case study, external representations play some contrasting of roles and do so by encoding information which plays different roles within the the reasoning process. The use of external working aids extends this variety. One way of looking at these roles is by associating the different kinds of information with elements within an argumentation structure. To this end we find a correspondence with three of the major elements in Toulmin's model of practical arguments (Toulmin, 1958). In Toulmin's model arguments are based on *data*. Data in the scenario is encoded as cues within TTO reports. It also appears in the form of results of prior inferences. Arguments result in inferred information or *conclusions*. We see these in the conclusions about level of command, unit identity and action. A third important element in Toulmin's model is the idea of a *warrant*. A warrant is a rule-like proposition which legitimises the inference from the data to conclusion, and in virtue of which an inference is possible. In our scenario, the role of the tables and the ORBAT is to provide those warrants through what we refer to as *mediating information*. They provide the basis on which the sensemaker can find meaning in cues. We show these three roles in Table 1, and list the information/artifacts which play those roles in the scenario.

Table 1. Cues (left), enabled information to be inferred (right) given mediating representations (middle).

Cue/information	**Mediating representation**	**Types of inferred information**
Frequency	Radio Equipment table	Level of command
Callsign	Radio Procedures – Callsigns table	
Type of encryption	ROM Encryption Systems Table	
Level of command, codewords, callsigns and message extracts	ORBAT	Unit identity
	Background knowledge	Action

Of particular interest in this study is the analyst's use of representational artefacts containing mediating information or warrants. In many sensemaking activities, warrant information is stored 'in the head' as background knowledge. This knowledge, which may be associative or rule-like, is what we associate with experience and drawn from long-term memory. In the case study the associations or rules are externalised and embodied in tables and charts. This is not to say that the analysts may not come to remember this information - in fact our experience with them showed that often they do, but knowledge represented in these tables and charts changes the nature of the activity, allowing the analyst to go beyond a potentially incomplete knowledge to increase the diagnostic reliability of any of a set of possibilities being considered.

Feltovich et al (1984) performed a study of medical diagnostic reasoning in which they referred to such a set of plausible possibilities as a Logical Competitor Set (LCS). In the study, medical students, trainees and experts were given case files and asked to make diagnoses, articulating hunches as they went. The study found a relationship between the number of items within the set that participants considered and their level of experience, concluding this to be a mark of expertise.

In the current study, Logical Competitor Sets with associated properties from which rules of inference could be constructed were externalised in the form of tables. The tables were printed on pieces of paper and the analysts had pens. The performance of each inference was a question of inspecting a range statement written in numerical characters (e.g. "1.25 - 4.5 MHz") to test whether a frequency fitted within the range. Recording the results of each inference was a question of whether or not to strike through with a pen a row in the Radio Equipment Table. And the result was a question of visually assessing the rows which were left. The table therefore serves as an external representation of judgements to be made about frequency, and as a memory of which possibilities have been eliminated, and which are still being considered.

The properties and affordances of representational artefacts are central to the way that they play a part in a distributed cognitive system. A distinction made by Hutchins that is illuminating in this case study is between the descriptive level of computational function, and the level of representation and implementation. At the computational level, the analysis system (of analysts and representational artefacts) makes inferences, constructs Logical Competitor Sets, and so on, along the way to producing the situation picture to be delivered to the supervisor. At the level of representation and implementation, the system can be one that manipulates, transforms, and combines representational media (for instance, by using information in a TTO to strike out items in the Radio Equipment Table, to narrow down the set of possible organisational units involved in a communication).

Inferences in sensemaking have been noted as often being abductive in character (Klein et al., 2007). Abduction, however is fallible since there may be plausible explanations missed. The process of elimination that we observed supported by material artefacts was, however, characteristically deductive. Deduction is truth preserving insofar as the result of a valid deduction necessarily true so long as the premises are true. So long as the reported frequency was correct and the information in the table was also correct and exhaustive, the conclusion would be guaranteed. Further, this deductive computation became possible in virtue of the material properties and affordances of the artefacts themselves. Our interpretation of the analysts strategies were that, in the face of the computation they wanted to perform and the material properties and affordances of the representational artefacts they were able to construct a strategy through which the computation could be implemented. Further, we assume that it is through 'seeing' the strategy as a possibility, as having a given user-cost and as providing an outcome which is helpful to the overall sensemaking task, the computation that it implements becomes a possibility and something worth doing. By changing the nature of the artefact, how it represents its information and how the user can interact with it, we might change any of the former properties and hence the properties of the associated computation.

In this paper we have explored, through an empirical study, how the performance of a military analysis task can be analysed as being an instance of Distributed Sensemaking. A central commitment of this analytic perspective is, following Hutchins' cognitive ethnographic approach, to take as an appropriate unit of analysis, not the actions or mental process of an individual analyst. Rather the unit of analysis used here is a distributed system of people and artefacts, and involves internal and external representational state and media. The analysis focuses on the properties and affordances of representational artefacts, such as the ease with which paper tables can be annotated in the computational process of making inferences. In this view, distributed sensemaking is seen as a process of transforming and propagating representational state in order to make interpretations, consider 'competitor sets', alternative hypotheses or 'frames', and develop a rich and reliable situation picture.

ACKNOWLEDGMENTS

The work reported in this paper was funded by the UK Defence Sciecne and Technology Laboratory via BAE Systems Defence Human Capability Science and Technology Centre as part of the Humans in Systems Programme. The training exercise was devised and run by MASS Consultants Ltd. and we would like to extend our gratitude to them for providing the research team with the opportunity to conduct the study.

REFERENCES

Dervin, B. (1983) An Overview of Sense-making Research: Concepts, Methods, and Results to date. Paper presented at the *International Communications Association Annual Meeting*, Dallas, May 1983.

Feltovich, P.J., Johnson, P.E., Moller, J.H., and Swanson, D.B. (1984). LCS: The role and development of medical knowledge in diagnostic expertise. In W.J. Clancey & E.H. Shortliffe (Eds.), *Readings in medical artificial intelligence: The first decade* (pp. 275-319). Reading, MA: Addison Wesley.

Furniss, D., & Blandford, A. (2006). Understanding Emergency Medical Dispatch in terms of Distributed Cognition: a case study. *Ergonomics*, *49*(12-13), 1174–1203.

Halverson, C.A. (1995) *Inside the Cognitive Workplace: New Technology and Air Traffic Control.* PhD Thesis, Dept. of Cognitive Science, University of California San Diego.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction*, *7*(2), 174–196.

Hutchins, E. (1995a). *Cognition in the Wild.* MIT press Cambridge, MA.

Hutchins, E. (1995b). How a cockpit remembers its speeds. *Cognitive Science*, *19*(3), 265–288.

Hutchins, E., & Klausen, T. (1998). Distributed cognition in an airline cockpit. In Engeström, Y., Middleton, D. (Eds) *Cognition and Communication at Work*, 15–34. CUP.

Klein, G., Phillips, J.K., Rall, E. & Peluso, D.A., (2007) A Data-Frame Theory of Sensemaking. In Hoffman R (Ed) *Expertise Out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making* (pp.113-155). New York, NY & Abingdon, UK: Lawrence Erlbaum Associates.

P. Pirolli and S. Card, (2005) The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis, Proceedings of the 2005 International Conference on Intelligence Analysis.

Rogers, Y. (2012). HCI theory: classical, modern, and contemporary. *Synthesis Lectures on Human-Centered Informatics*, *5*(2), 1–129.

Russell, D. M., Stefik, M. J., Pirolli, P. and Card, S. K. (1993) The Cost Structure of Sensemaking. In Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (pp. 269-276), ACM Press, New York, NY, pp. 269-276.

Toulmin, S. (1958). *The Uses of Argument*, Cambridge, England: Cambridge University Press.

Weick, K. *Sensemaking in Organisations*. Sage, London, England, 1995.

Wright, P. C., Fields, R. E., & Harrison, M. D. (2000). Analyzing human-computer interaction as distributed cognition: the resources model. *Human-Computer Interaction*, *15*(1), 1–41.

Evidence-Based Decision Making in Civilian Agencies: An Analysis of Three Cases

Suzanne L. GEIGLE
MITRE Corporation

ABSTRACT

Evidence-based decision making has emerged as an area of needed improvement in government. A brief overview of the research on individual decision making suggests why managers in both the public and private sectors have failed to take advantage of the best available evidence when making decisions. Enter the evidence-based management movement. Borrowing lessons from evidence-based medicine, management scholars and practitioners have formed a community focused on developing methods and tools to help practitioners seek and use the best available evidence in decision making, which in turn will improve decision outcomes. This study contributes to the discussion in that community by offering case studies focused on the current use of evidence in decision making at three civilian agencies. Relevant models that guided data collection and analysis are presented. Findings suggest the main barriers to adoption of evidence-based management practices and how external advisors can play a key role in achieving improvements.

KEYWORDS

Decision Making, Civilian Government, Use of Evidence, Stakeholders, Strategic Advisors

INTRODUCTION

Today, studies of actual decision making have led management scholars to conclude that a theory of decision making based entirely on rational choice is of limited usefulness. In practice, managers' decisions are made without awareness of all alternatives and preferences, without considering all consequences, and often without a clear sense of the goals to be achieved. Decision makers tend to focus on some types of information and ignore others and follow decision rules that vary from one situation to another. These studies support the concept of limited or bounded rationality (March, 1994), which has come to dominate most theories of individual decision making. The concept assumes that individuals try to be rational decision makers, but they are constrained by limited cognitive capabilities and incomplete information. As a result, their actions may be less than rational in spite of their best efforts. March's work on limited rationality and the tactics employed by decision makers to simplify a complex situation or problem highlights how these common and largely automatic simplification activities often lead to errors.

In government organizations, the impact of errors in decision making by government managers and executives can have far-reaching effects on their programs, funding decisions, and ultimately on the citizens they serve. Concerns about the quality of decisions in government have prompted an increased focus on "data-driven" or "evidence-based" decision making practices. For example, in his May 18, 2012 memo to the heads of executive departments and agencies, OMB Acting Director Jeffrey Zients strongly encouraged agencies to "demonstrate the use of evidence throughout their Fiscal Year (FY) 2014 budget submissions" (M-12-14). His specific guidance drew attention to the important role that evidence would play in the evaluation of budget submissions and suggested OMB's awareness that evidence-based decision making in government was the exception rather than the rule.

When decision makers are told to support their decisions with evidence, most are likely to think they already do this. Yet, observations of government and private sector decision makers (Pfeffer & Sutton, 2006) suggest that leaders' decision making practices are driven more by personal experience than by a systematic scan of relevant evidence. The practices in use are more consistent with a search for an action or solution that is good enough, rather than a drive toward the best possible solution (March, 1994). In addition, changing the customary decision making practices of leaders is particularly challenging and likely to trigger strong resistance (Yates, 2003). In summary, decision makers of all stripes tend to be overconfident in their knowledge of relevant facts and often unable to recognize the impact that cognitive biases have on the quality of their decisions. In situations where the problems that leaders face are new or are poorly understood, these cognitive threats to effective decision making are heightened. Enter evidence-based decision making, which is thought to help decision makers counteract the errors that tend to occur when problem simplification activities, i.e., editing, decomposition, heuristics, and framing (March, 1994), are invoked during decision making.

For more than a decade, management educators in the US and abroad have raised concerns about a related problem -- the failure of management science to bridge the research-practice gap (Rousseau, 2006). Their concern about bridging the research-practice gap has helped spawn a movement by scholars, educators, and practitioners, which is intended to overcome managers' unwillingness or inability to leverage the best evidence available and thus mitigate the effects of limited cognition in decision making. This group of scholars and practitioners continues to address the research-practice gap under the umbrella of evidence-based management (http://www.cebma.org, Rousseau, 2012). Drawing lessons from the adoption of evidence-based medicine by medical practitioners, advocates of evidence-based management actively seek methods and tools that will help management practitioners adopt decision-making practices that improve the quality of decisions made and actions taken.

This study, when originally conceived, was intended to explore why evidence-based decision making in civilian government organizations is so rare. The choice to focus on civilian agencies was based on the assumption that too little is known about the drivers of decision making outside the military, where formal decision making processes are the norm (citation?). If our ultimate goal as strategic advisors is to help agency managers seek and use the right evidence at the right time during decision making, knowledge of when and how they use evidence now (if at all) is essential to tailoring an intervention that would produce better decision outcomes.

Most studies of decision making focus on decision processes. This study, in contrast, focused primarily on the use of evidence in whatever decision process was underway. We wanted to know the following: when agencies are faced with important decisions, what kinds of evidence do they seek, and how and when do they use evidence to support those decisions?

METHODS

The plan for the study was to conduct an analysis of several decisions, preferably in different civilian agencies. This called for a case study approach, which would allow the researcher to explore in depth the activity in question, gaining perspectives from multiple individuals involved in the activity. Each case is bounded by time and activity (Creswell, 2009). The primary data collection methods used were participant observation and interviews. The researcher also requested documents associated with the decision problem, which were thought to suggest where evidence had been offered.

Three research sites were identified for the case studies. All three were civilian government agencies that were making significant decisions affecting their organizations and the customers they served. The original design called for the researcher to observe how evidence is sought and used during decision making in real time. However, finding willing research sites was a greater challenge than expected, which forced the researcher to study a set of cases that met only some of the original criteria for selection. In the end, none of the cases studied and reported here fit the desired longitudinal design. In addition, each case presented a different set of data collection constraints, which are described in the Findings section.

A Model of Evidence Sources and Practices

Data collection and analysis were guided using a model of evidence-based decision making developed by scholars studying and teaching evidence-based management. Their model, adapted by the researcher in Figure 1, posited that evidence-based management would follow from the conscientious, explicit, and judicious use of four sources of information: practitioner expertise and judgment, evidence from the local context, a critical evaluation of the best available research evidence, and the perspectives of those people who might be affected by the decision (Briner, Denyer, & Rousseau, 2009).

Figure 1: Sources that Support Evidence-based Management (adapted from Briner, et al)

In February 2012, one of the model's authors addressed an audience of systems engineers at MITRE Corporation, where she made explicit the relationship between evidence-based management and decision making. Comments from this speech are reproduced in Figure 2.

> "Evidence-based management is the practice of making organizational decisions based upon
>
> - Conscientious use of science-based principles,
> - Valid and reliable organizational facts,
> - Decision supports and reflective judgment, and
> - Ethical considerations, particularly as related to stakeholders.
>
> The result is improved decision quality through more consistent use of practices that work" (Rousseau, 2012a).

Figure 2: Practices that Support Evidence-based Management

The interview protocol was then designed to gather data on both types of evidence and practices used in order to determine when and how evidence played a role in the decision making process. (Interview guide available on request.) Observation notes and interview text were coded using NVivo as part of the analysis.

FINDINGS

This section presents a short description of each case followed by the results of its within-case analysis. In order to illustrate how the model of evidence-based decision making was used for within-in case analysis, the tables generated during analysis of Case 3 data are included (Tables 1 & 2). After presenting the three cases, the section concludes the results of a cross-case analysis.

Case One: Customer Service Plan (CSP). The decision studied in the CSP case was triggered by the following events: an agency needed to establish a new customer service capability in response to a new legislative mandate. The capability had to be up and running in less than a year despite competing demands for resources. For the first time in the agency's history, it would have to work with another agency in another Cabinet department to be successful. Managers and executives from different parts of the organization saw the challenge differently, each through his or her own frame of reference. The accountable decision makers took unusual actions in this case. The decision outcomes were perceived as positive because the agency met its deadline. Data collected in this case came from one key informant, the deputy project manager (DPM) who witnessed the decision making process unfold but who was not a member of the executive decision making group. The researcher conducted two two-hour interviews with the DPM to collect data on the relevant events.

The data available to the researcher on CSP were limited to one person's perspective and were based on recollections after the fact. Given these limitations, two observations are still worth noting. First, top executives at the agency allowed the researcher access to the DPM -- despite the heavy work demands -- because this case of decision making was viewed as highly successful. Second, when describing the activities that made this experience successful, the DPM pointed immediately to early engagement with key stakeholders. She was referring to a "road show" conducted by the executive decision makers, which consisted of some education about the challenges the agency faced, some opportunities for stakeholders to express concerns and questions, and finally, an invitation to participate in the solution.

Case Two: Reorganization (REORG). This case followed the activities of an executive team formed to plan and implement a major reorganization of the agency. The researcher analysed documents and observed two all-day working sessions of this executive team, capturing the discussion as close to verbatim as possible. Observation notes were coded to identify what evidence was used and how it was used during those working sessions. A series of follow-up individual interviews with members of the executive team provided insights into members' thoughts about their decision making process and the drivers of the group's decisions. In this case, several FFRDC advisors were contracted to conduct an employee survey; this work was completed prior to the start of this research study. The FFRDC advisors were interviewed for this study to ascertain when and how the data they gathered from employees had been used by the REORG team.

This case study analysis was based on multiple sources of data. A within-case analysis of REORG produced the following findings:

- **Process guidance came from an oversight committee.** A formal process was used to guide activities. This guidance came from a federal oversight committee's highly critical report on previous reorganizations of the agency. The REORG team chairperson returned to this report frequently to assess progress.
- **The impact of employee (stakeholder) survey results on decisions was unclear.** While the process guide advised them to collect employee data, the executive team struggled to determine how to respond to and use the employee feedback from the survey. At the first working session observed by the researcher, the executive team devoted more than an hour to this topic. Several months had passed since the survey was administered and the results delivered, yet the executive team still had not conducted the promised all-hands meeting to report results. Team members expressed concern that the delay in feeding back the results of the employee survey was undermining the credibility of the REORG process and the team. In addition, the content of employee survey results (stakeholder preferences) was never mentioned during the observed working sessions.
- **Organizational facts were presented, but did not move the decision process forward.** Sub-teams of the REORG team worked on specific issues between meetings of the full team. At all-day working sessions, sub-teams presented their findings, which generated extended discussions. In the two sessions observed, no decisions were teed up by the chair. In one case, a decision agreed to at a previous working session was revisited for discussion and reconsideration. The effect of these infusions of new evidence appeared to work against reaching interim decisions that would stick.
- **The leader's experience and decision style dominated discussions.** The leader of the REORG had decades of experience at this agency. His stories about what happened in the past consumed a good bit of air time at the observed working sessions. Other members of the REORG team tolerated these digressions. The researcher did not observe any activities during working sessions that brought discussions to closure and decisions; this appeared to reflect the leader's meeting management style. In follow-up interviews, team members acknowledged that the leader dominated working sessions and tended to resist final decisions, but did not express any concern about its effect on the group's overall decision process.
- **External evidence rarely used.** Apart from the oversight committee's report, the REORG team did not seek outside guidance on reorganizations. The FFRDC advisors who had conducted the employee survey offered to provide subject matter expert (SME) support, but this was rejected by the REORG team leader. Observation notes from working sessions pinpoint several discussions where an injection of "best available" evidence from an external source would have been helpful. It was clear that this knowledge did not exist within the team. These "teachable moments" passed without the benefit of the evaluated external evidence noted in the model in Figure 1.

Case Three: Strategic Goals. The driver of the Strategic Goals case was the appointment of a new CIO for a civilian agency and his desire to set a new strategic direction. He tasked one of his division chiefs to lead a project that would develop a new set of strategic goals for this enterprise-wide IT department. When the researcher became aware of this case, the project had already ended several months earlier. A new set of strategic goals had been developed and approved by relevant governing bodies. While implementation planning had just started, the process used to set new goals was viewed within the agency as a great success. The researcher started by interviewing the FFRDC advisor who had worked with the decision makers; this provided a description of the key steps the department leaders had followed. Next, the researcher interviewed the division chief who led the process. Following the interview protocol, this key informant offered data on what evidence was sought and used at each step. Tables 1 and 2 show the within-case analysis of sources and practices in the Strategic Goals case.

Table 1 shows how the data collected in this case mapped to the sources of evidence described in Figure 1. Table 2 shows the specific practices used in this case, mapped to the evidence-based decision making practices described in Figure 2.

Table 1. Sources of evidence in the Strategic Goals case

Evidence-based DM Sources (Figure 1)	Analysis of Strategic Goals Case Data
Practitioner experience and judgment	CIO's prior experience led him to focus on three objectives: 1) Pull together plan to accomplish things; 2) Get the management team functioning more effectively; 3) Foster a sense of community across the enterprise to help make decision making more efficient. CIO adjusted his plans over the year as more data were collected; this showed flexibility and a realistic sense of what could be accomplished and when.
Context, organizational actors, circumstance	CIO's experience made him sensitive to the decentralized governance structure in this organization; he knew he would need to use a different approach than he had used at other organizations; he already knew who the key actors were in the organization and whose support would be needed to go forward with his plan A review of previous strategy documents revealed little of use -- "a bit of marketing fluff" -- but was an important first step First strategic planning meeting was held with his own staff, which helped him to get a handle on his own shop first and to assess how the resources under his direct control could be used more effectively
Stakeholders' preferences or values	Conducted off-site meetings with execs from all parts of the organization to listen to their views and needs; validated all draft plans with this group Gathered data from his direct reports on current priorities, thus demonstrating that their views are important
Evaluated external evidence	Used a commercial tool, developed more than 10 years earlier and used widely in other organizations, to collect data on each project; this enforced a greater level of consistency in the data for decision making

Table 2. Evidence-based practices in the Strategic Goals case

Evidence-based DM Practices (Figure 2)	Analysis of Strategic Goals Case Data
Conscientious use of science-based principles	Face-to-face meetings used to gather input from stakeholders. [Research has shown that face-to-face communication is the richest medium for information exchange because it is immediate and personal, and includes auditory, visual, and non-verbal behavioral data. Face-to-face communication thus reduces ambiguity and confusion. (Daft & Lengel, 1986)] Standard commercial business case template used to gather complete and consistent data across all projects
Valid and reliable organizational facts	Conducted first-hand research on stakeholder needs; this ensured that the information used to make decisions was current and reduced the chances of misunderstanding stakeholder comments Used FFRDC to ensure that stakeholder data was gathered and managed by an objective third party, which contributes to the credibility of the process and information Representation from all organizational components improves the validity of data collected
Decision	Commercial tool organizes data from different projects in a consistent

supports and reflective judgment	way, which helps decision makers make comparisons and draw valid conclusions Negotiated solutions were needed in some cases in order to reach agreement among all key stakeholders, aka the "collegiality tax"; this diversity of opinions forced the Strategic Goals team to develop a more innovative approach than their obvious first choice
Ethical considerations, particularly related to stakeholders	The Strategic Goals team recognized the need for socializing new ideas and being open to solutions offered by others. This approach to issue resolution, i.e., giving stakeholders a voice in a decision-making process, is perceived as more fair.

Cross-case Analysis

Potential research sites for decision making studies are most likely to be "success stories." These three agencies viewed the decision making activities they used as successful. This factor emerged as most important during site selection, because it determined whether the organizations were willing to share what happened in the past and what was occurring in the present. Decision making can be messy and leaders are often reluctant to shine a light on less-than-positive features of the organization and its routines. In addition, organizations are understandably reticent about sharing what usually goes on behind closed doors. In these three cases, the organizations were open to the research because they were proud of the outcomes or they believed they were using a good process.

Stakeholder outreach was a major factor and *unusual*. In all three cases, decision makers made a deliberate effort to gather data on stakeholder preferences, needs, concerns, etc. In all three cases, informants expressed how unusual this outreach to stakeholders was in their organizations. Informants in two of the three cases attributed the success of the decision process specifically to stakeholder outreach. This suggests that when the action of engaging stakeholders is unusual, it fosters support for the outcomes by both the decision makers and those affected. In two of the three cases, (CSP and Strategic Goals), the researcher was able to ascertain the positive effect of evidence gathered from stakeholders on final decisions. In the REORG case, it is possible that the decision makers' inability to make productive use of evidence from stakeholders during deliberations worked against them. During a follow-up discussion with those familiar with the case, the researcher learned that decisions made by this REORG team were not implemented.

External "best available" evidence was rarely sought out. In all three cases, decision makers relied most on personal experience and knowledge of the organizational context to guide their decisions. In only one case (Strategic Goal), the leader brought into the discussion his experiences from previous employers. In this case, the FFRDC provided intermittent facilitation support to the decision makers, but the agency did not explicitly contract with the FFRDC to bring external evidence into discussion.

DISCUSSION

These findings suggest that civilian government agencies may find it most difficult to overcome the natural cognitive bias associated with personal experience. They also may be reluctant to include outsiders in decision processes that have been traditionally restricted to executives and may have been shrouded in secrecy. However, outside advisors could provide a valuable service -- seeking out and injecting relevant external evidence at teachable moments.

The importance of teachable moments during decision making became clear in the analysis of two of the case studies. In the REORG case, because no knowledgeable external SME was permitted into the working sessions, the group relied completely on individuals' past experience in their own organization. This limited their ability to see alternatives. They did not know there was pertinent evidence available that could shed light on the problem they were discussing. In short, they did not know what they did not know. In the Strategic Goals case, the participation of an external advisor (FFRDC practitioner) during planning and execution of working sessions allowed the decision makers to invite and entertain new ideas. The external advisor was able to suggest bringing in external SMEs when appropriate, which he did on at least one occasion. In this case, the external advisor was not pushing a proprietary methodology – a criticism of many consultancies – but instead listened for the group's needs before offering suggestions. In this case, the advisor stood alongside the team, offering guidance when asked and making timely suggestions when an opening for a new idea and new evidence appeared.

CONCLUSION

These findings make no assertions of generalizability. On the contrary, the study is subject to the following limitations: 1) only three cases could be analysed because examples of decision making processes in civilian agencies took months to identify and to negotiate entry; 2) the processes studied could not be followed in real time as the research period was too limited; and 3) data collection was constrained by restrictions on access to agency employees. For these reasons, the findings are considered suggestive.

The Role of the External Advisor in Evidence-based Decision Making

The most fruitful path to gaining greater adoption of evidence-based decision making practices will be through external advisors, at least in the short-term. External advisors are less constrained by internal organizational politics, so they should be able to provide better quality evidence to decision makers. However, such an advisor's success will depend on having established trusted relationships with decision makers, which would, in turn, permit her to be present at teachable moments when decisions are in development. The success of the advisor also depends on her ability to recognize when and what kind of evidence is needed and her ability to reach out to a network of SMEs (and a relevant body of evidence) when needed. The challenge of serving in this role deserves focused attention by organizations like FFRDCs, which seek to be trusted advisors to their government sponsors.

Evidence-based practices are especially important to organizational practitioners who serve in advisory roles, those who "in many instances are not the key decision makers but, rather, sources of information and advice to managers making the decision." FFRDC practitioners, who are often in such roles, are thus well-positioned to work "as facilitators and coaches for managers and management teams seeking to engage in evidence-based management, as well as helping them to collect internal and external evidence they may need" (Briner & Rousseau, 2011, p. 20). However, even the most trusted advisor to decision makers cannot be present for every teachable moment. This reality suggests the need for strategic advisors to develop more ways of bringing evidence to bear on practice. Two suggestions offered by Briner and Rousseau are the development of practice-oriented evidence and systematic reviews.

In summary, for decision makers in organizations to get the benefits of evidence-based practices, they will often need the active intervention of a skilled external advisor/facilitator/coach. Because evidence-based decision making relies first and foremost on assembling meaningful, reliable internal and external evidence for careful consideration, the role of the external advisor is different from a smart consultant who confuses opinions with evidence. Instead, the external advisor can have significant impact when we adhere conscientiously to science-based principles when vetting external information sources, and employ rigorous data collection and analysis methods when gathering internal information. That is what it takes to bring quality evidence to decision making.

REFERENCES

Briner, R. B., Denyer, D., & Rousseau, D. (2009). Evidence-based management: Concept cleanup time? *Academy of Management Perspectives*, 23(4), 19-32.

Briner, R. B., & Rousseau, D. (2011). Evidence-Based I-O psychology: Not there yet. *Industrial and Organizational Psychology-Perspectives on Science and Practice*, 4(1), 3-22.

Creswell, J. W. (2009). *Research Design* 3d ed. Thousand Oaks, CA: Sage Publications.

Daft, R. L., & Lengel, R. H. (1986). Organization information requirements, media richness, and structural design. *Management Science*, 32(5), 554-571.

March, J. G. (1994). *A Primer on decision making: How decisions happen*. New York: The Free Press.

Pfeffer, J., & Sutton, R. I. (2006). *Hard facts, dangerous half-truths, and total nonsense*. Boston: Harvard Business School Press.

Rousseau, D. M. (2006). Is there such a thing as "evidence-based management"? *Academy of Management Review*, 31, 256-269.

Rousseau, D. M., ed. (2012). *Oxford Handbook of Evidence-based Management*. Oxford University Press.

Rousseau, D. M. (2012a). Evidence-based Decision Making. *MITRE Technical Exchange Meeting speech delivered on February 16, 2012*. McLean, VA.

Yates, J. F. (2003). *Decision Management: How to Assure Better Decisions in Your Company*. San Francisco, CA: Jossey-Bass.

Intuitive Potential and Predicting Entrepreneurship – a Study on a New Method of Measuring Intuition

Matylda GERBER[a]
[a]*Warsaw School of Economics, University of Social Sciences and Hummanities*

ABSTRACT
Although some studies have been conducted to measure the preference to use the intuitive decision-making over the analytical one, there is still a high difficulty to estimate the real potential of intuition. A possibility to assess the potential of intuition would be crucial both for venture capitals and owners of organizations to elect future successful entrepreneurs and intrapreneurs from other candidates. This study is a proposition of a new way of measuring intuition to distinguish successful entrepreneurs. Using an authorial tool to measure the potential of intuition, the results indicated that entrepreneurs have higher potential than employees, but they do not differ in the preference for the intuitive decision-making over the analytical one. Even though the study was conducted on a small sample of 30 entrepreneurs and 30 employees from Poland, the results encourage a further research in this field of study.

KEYWORDS
Planning and Prediction, Business, Intuition, Insight, Entrepreneurship

INTRODUCTION

An entrepreneur is perceived as a creative person that modifies or rejects previously accapted ideas to build innovations. For these attributes an entrepreneur needs not only creativity but also intuition. The ability to gather new ideas from a nonconscious analysis of one's stored knowledge and experience would be crucial for successful entrepreneurs (Engle, Mah, Shardi, 1997). According to Sternberg (2004) entrepreneurship is connected with a mix of: general intelligence, intuition, creativity and analytical skills. However, intuition and creativity are perceived to be those cognitive factors that distinguish successful entrepreneurs (Baron, 1998, Kao,1989; Sexton&Bowman-Upton, 1991).

There have already been conducted many studies that measure entrepreneurial attitude by using questionnaries (Kirton, 1976; Buttner, Gryskiewicz, 1993; Allinson, Hayes, 1996; Engle, Mah, Sadri, 1997). However, there were only a few attempts to investigate a real entrepreneurial potential. Ames and Runco (2005) analyzed whether more successful entrepreneurs have a higher ideation potential than the less successful ones. They used a self-report measure and a SWOT analysis task. Unfortunately, the proposed task failed to be diagnostic, probably, because it was difficult to measure and to compare results and it was too time-consuming, which could discourage participants from getting fully involved. Not only Ames and Runco (2005) claim that it is very difficult to measure the entrepreneurial potential but also Blume and Covin (2011) admit this toughness, especially in the attempt to measure unconscious processes. It is because there is still a little knowledge about how unconscious data processing really works (Nosal, 2009). Intuition which is perecived as „affectively charged judgments that arise through rapid non-conscious and holistic associations", (Dane, Pratt 2007 p. 40) could be assesed looking at its result that appears in consciousness. Taking into account that we gather intuition through experiencing insights, we can estimate a potential of intuition by measuring the rapidity of attaining insights while solving new problems (Nosal, 2010). Insight, which is perceived as „sudden unexpected thoughts that solve problems" (Hogarth, 2001, p.251) was already measured using different tasks, which required from a person to redesign a given problem to find the solution (Sternberg, Davidson, 1996). Among different tools measuring the insight potential, there were considered Bongard problems to be a reliable way to distinguish people with higher insight potential from those with the lower one (Hofstadter, 1979, Tubek, Piskorz 1994). All things considered, we can assume that:

Hypothesis 1: Entrepreneurs have a higher insight potential (solve more Bongard problems) than employees.
Hypothesis 2: There is a high, positive correlation between the level of success and the level of insight (a number of solved Bongard problems) among entrepreneurs.

As described above, it is possible to distinguish entrepreneurs from employees using questionnaires. Allinson and Hayes (1996) using Cognitive Style Index showed that entrepreneurs are more intuitive than non-owner managers.

Furthermore, Engle et al (1997) proved, using Kirton's Adaptation-Innovation Inventory, that entrepreneurs are characterized by higher intuition than employees. Based on that research it could be assumed that entrepreneurs would have a higher score in the KSP questionnaire (Nosal, Sobków, 2012) on the intuition scale, where intuition is defined as: processing information in a holistic way, concentrating on general regulatories and making decisions based on a hunch and spontaneous learning patterns.

Intuition is strongly interrelated with a level of data processing (Kapur et al., 1994). Deep processing is responsible for successful retrieval of stored information (Craik 2002) and of processing general and abstractive data of an experienced event (Cohen, 2000; Conway, 1992). Based on that information, it should be assumed that entrepreneurs have not only higher score on the intuition scale but also on the depth of processing scale, which is defined as critical evaluation of existing information and goes beyond a given material (Nosal, Sobków, 2012). Based on presented information, the following hypothesis could be assumed :

Hypothesis 3: Entrepreneurs have a higher score on the intuition scale than employees.
Hypothesis 4: Entrepreneurs have a higher score on the depth of processing scale than employees.

Taking into account that there were only few studies concerning the applicability of intuition in entrepreneurial decision making, there is an ambiguity if individuals rely on intuition in their decision making or they just believe that intuition is informing their decision (Blume, Covin, 2011). According to Karwowski (2009) attitudes could really differ from the real potential. He proved that there is no relationship between creative attitude and creative potential. Based on Karwowski's (2009) research result it could be assumed that it would be similar with intuition. The preference to use intuition in one's decision-making could have no links with the real potential of intuition. It could be assumed that :

Hypothesis 5: There is no relationship between the level of insight (number of solved Bongard problems) and the score on the intuitive scale.

METHOD

Participants

30 entrepreneurs and 30 employees from Poland participated in the research (40 men and 20 women). All of them were from: Healt Care, IT or Banking sector and they had at least 1 year experience in a declared business. 15 entrepreneurs were recruted from the Polish Private Hospitals Association (OSSP) via emails or during a congress for the members of the association. The rest of research participants were postgraduate students from the University of Economics in Wroclaw and Warsaw School of Economics. The additional criterium for the group of employees was no intention to set up their own business.

Materials

The research tool consisted of 3 parts:

1. KSP Questionnaire (Nosal, Sobków, 2012) measuring intuition and the depth of processing,
2. Bongard problems that measure the insight potential,
3. Index of success measuring the level of success obtained by the run company.

The KSP Questionnaire includes 27 statements, where a subject has to decide on a 4-point Likert scale, to what extent he or she agrees with a given statement. The questionnaire consists of 2 scales: intuition and the depth of processing.

The second part of the research was desinged by the author of this paper. It consists of 11 diagnostic pictures and 2 samples (an example of pictures is presented in Figure 1). Those graphic tasks were chosen from a bank of 280 pictures, from a web page: http://www.foundalis.com/res/bps/bpidx/htm. 100 of them are the orginal idea of Bongard (1970) and the rest of them were created on the basis of the same idea. Their initial application was for a computer programme creation. For the pilot study, 27 graphic tasks were chosen, sufficiently complex to involve nonconscious processes to their solutions. 29 pilot study participants also solved 19 text tasks that measure insight potential. The study was conducted online, where the subjects were instructed to solve the gratest amount of tasks in the shortest time. Time was counted automaticly for each exercise. The analysis of the answers showed a high correlation between Bongard tasks and text tasks solved ($r=,641$, $p<0,01$), which poses an additional proof that those picture tasks measure the insight. Taking into account the fact that tested Bongard problems turned out to be very difficult, 11 tasks that were the most frequently solved in the shortest period of time, were chosen.

Additionaly, because of many complaints form the study participants that the inability to solve tasks was very frustrating, the information, that some of the tasks could have no solution and the correct answer would be „lack of rule", was added.

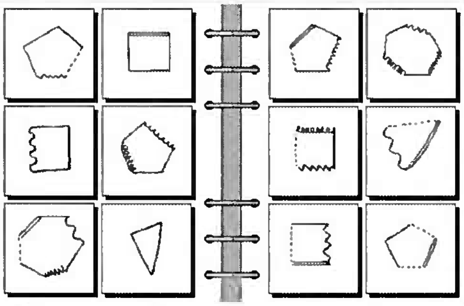


Figure 1: an exaple of Bongard problem used in the research.

The third part of the research consisted of 4 indicators that measure the level of success achieved by companies, where a person had to specify:

- Whether the increase of sales was higher, lower or similar to the market average,
- Whether the managed enterprise achieved an increase of profitability, comparing to previous years,
- Whether there was an improvement in customers' opinion in comparison to previous years,
- Whether the company achieved one of the specified business awards.

This part of the research covered only a group of entrepreneurs. Those indicators were consulted with a partner of EY consulting company – a person with an extensive experience in assesing success of enterprises.

Procedure

Participants received an e-mail with: the study decription, an assurance of anonimity of participation, instructions and a link to the online research. The first task was to score to what extent a person agrees with given 27 statements, that come from KSP questionnarie (Nosal, Sobków 2012). After completing this part of the study, they were acquainted with the instructions how to solve graphic tasks. They had to find and to describe the rule that differentiates 6 pictures on the left from the 6 ones on the right or to write „lack of rule", when no rule could be found. At the beginning there were 2 example tasks with correct answers presented and later on, one after another, 11 diagnostic tasks appeared. When this part was finnished, for those participants who were entrepreneurs, the last part with 4 questions measuring the level of success of their company was presented.

RESULTS

To test Hypothesis 1, stating that entrepreneurs have a higher insight potential (solve more Bongard problems) than employees, Mann-Whitney U test was used, because distribution in both the entrepreneurs and employees groups differed significantly from the normal one. The comparision indicated significant differences: $U=310{,}00$ ($p<0.05$) in the number of solved Bongard problems. It was significantly higher in the group of entrepreneurs (avg=4,2) than in the group of employees (avg=2,9). This result confirms Hypothesis 1.

Due to significant discrepancies of the analysed variables with the normal distribution, to test Hypothesis 2, the Spearman rho correlation coefficient was calculated. This analysis considered only the group of entrepreneurs, because only in this group the index of success was measured. The analysis indicated a high, positive and significant correlation between variables: rho=0,698 ($p<0{,}001$). This result confirms Hypothesis 2, that there is a high, positive correlation between the level of success and the level of insight (a number of solved Bongard problems) among entrepreneurs.

Hypotheses 3 and 4, stating that entrepreneurs have a higher score on the intuition and on the depth of processing scales than employees were tested using the t-Student test. The results are presented in Table 1.

		M	SD	t-Student	p
intuition	entrepreneur	3,22	,33	t(58)=0,86	0,39
	employee	3,14	,41		
Depth of processing	entrepreneur	3,39	,33	t(58)=1,23	0,22
	employee	3,28	,37		

Table 1 t-Student test comparison of intuition variable in groups: entrepreneurs and employees

The comparison did not indicate any significant differences between entrepreneurs and employees among variables: intuition and the depth of processing, what means that Hypothesis 3 and 4 were not confirmed.

Taking into account a significant discrepancy between the analysed variances with the normal distribution, the Spearman rho to test Hypothesis 5 was counted. The analysis did not indicate any significant correlation (rho=,101, p=0,44), what confirms Hypothesis 5, that there is no relationship between a number of solved Bongard problems and the score on the intuition scale.

DISCUSSION

The results of this study indicate that entrepreneurs have a higher potential of intuition than employees and that it is possible to measure this potential. It also shows that those entrepreneurs who are more successful in running their business have also better intuition. However, the research has not proved that entrepreneurs prefer to make decisions more intuitively than employees and that they process information deeper than the later ones. Finally, the study confirms that a preference for intuitive decision-making over analytical one does not equate with the real potential of intuition.

The big advantage of the research is that there was used a self-designed tool to measure the insight potential and that it was conducted on a group of entrepreneurs and employees - which is difficult to acquire but very important from scientific and business point of view. Even though Bongard problems were already considered to be a good measure of the insight (Hofstadter, 1979; Tubck, Piskorz, 1994), it was the first attempt to use it as a tool to diagnose entrepreneurial potential.

The second aspect, important to notice, is that there was no correlation between the potential of intuition and a preference for intuitive decision-making. It is consistent with Karwowski's research (2009) on creativity, which shows that people could be very creative and have a creative attitude but also those with creative attitude could not be very creative. Moreover, it happens also that both very creative and less creative ones could not have this attitude. This presents the importance of measuring the real potential of intuition, not just a preference.

The surprising results showing that there is no difference in the level of intuition and the depth of processing among entrepreneurs and employees could be caused by the sampling. People from both groups were very similar, they were from the same business sectors and with at least 1 year experience. Another reason could be seen in the measuring tool. Despite the fact that the KSP questionnaire (Nosal, Sobków, 2012) is very reliable; Cronbach's alpha for intuition is 0,722 and for depth of processing 0,833, the questionnaire was not validated.

Despite these contributions, this study is not without limitations. It has a highly experimental form. Even though results of the research gave promising information, it should be tested on a much bigger group of participants. Moreover, the concept of Bongard problems is to measure the insight potential, which is not identical with intuition. According to Dane and Pratt (2007) one of the biggest differences between insight and intuition is that in the first case we are able to consciously become aware of the logic that has led to the result, whereas in the second phenomenon it is rather not possible. Moreover, the utilized tasks to measure the insight have only one correct answer, what means that to solve them the convergent thinking is needed (Sternberg, Davidson, 1996). However, intuition is perceived to be more strongly connected with the divergent one (Ames, Runco, 2005). Another drawback of the research concerns the index of success. Participants of the research were asked to assess subjectively the results of a company. Such results are never fully reliable.

To sum up, the presented research should be treated as an attempt to create a reliable tool to investigate the potential of intuition, which would be very important both for venture capitals – to test whether investing in given entrepreneurs might be successful and for owners of organizations – to recruit and to select those employees who would highly contribute to the development of an organization. It is also important to notice that even though it is easier to measure a given phenomenon using a questionnaire, utilizing tools that assess the real potential give the most reliable results.

ACKNOWLEDGMENTS

I would liket to thank: Zbigniew Piskorz, the OSSP organization and their members, Lukasz Zalicki and Jakub Szalaty for the substantive support while creating and conducting the research.

I would like to thank also Zbigniew Piskorz, PhD and Czeslaw Nosal, full Professor, for recommending Bongard problems as a tool to measure insight.

REFERENCES

Ames, M., Runco, M. A. (2005). Predicting entrepreneurship from ideation and divergent thinking. *Creativity and Innovation Management, 14*, 311-315.

Baron, R. (1998). Cognitive mechanisms in entrepreneurship: Why and when entrepreneurs think differently than other people. *Journal of Business Venturing, 12,* 275-94.

Baron, R. (2007). Behavioral and cognitive factors in entrepreneurship: Entrepreneurs as the active element in new venture creation. *Strategic Entrepreneurship Journal*, 1, 167-182.

Blume, D., Covin, J. (2011). Attributions to intuition in the venture founding process: Do entrepreneurs actually use intuition or just say that they do? *Journal of Business Venturing.*

Bongard, M. M. (1970). Pattern Recognition. Rochelle Park, N.J.: *Hayden Book Co.*

Buttner, E. H., Gryskiewicz, N. (1993). Entrepreneurs' problem-solving styles: an empirical study using the Kirton Adaption/Innovation theory. *Journal of Small Business Management*, 1, 22-31.

Cohen, G. (2000). Hierarchical models in cognition: Do they have psychological reality? *European Journal of Cognitive Psychology*, *12*, 1–36.

Conway, M.A. (1992). A structural model of autobiographical memory. In M.A. Conway, D.C. Rubin, H. Spinner, W.A. Wagenaar (Ees.), *Theoretical perspectives on autobiographical memory* (pp. 167-193). Dordrecht: Kulwer Academic.

Craik, F. I. M. (2002). Levels of processing: Past, present and future? *Memory*, *10*(5/6), 305-318.

Engle,.D., Mah, J., Sadri, G. (1997). An empirical Comparison of Entrepreneurs and Employees: Implications for Innovation. *Creativity Research Journal.*

Hofstadter, D. (1979). Gödel, Escher, Bach: An Eternal Golden Braid. *Basic Books.*

Hogarth, R. (2001). Eudcating Intuition. Chicago: *University of Chicago Press.*

Kao, J. (1989). Entrepreneurship, reativity, and organizations. *Prentence Hall.*

Kapur, S., Craik, F. I. M, Tulving, E., Wilson, A. A., Houle, S., & Brown, G. M. (1994). Neuroanatomical correlates of encoding in episodic memory: Levels of processing effect. *Proceedings of the National Academy of Sciences of the United States of America, 91*, 2008-2011.

Karwowski, M. (2009). Zgłębianie kreatywności. Studia nad pomiarem poziomu i stylu twórczości. *Wydawnictwo APS.*

Kirton, M. (1976). Adaptor and innovators: A description and measure. *Journal of Applied Psychology,* 61, 622-629.

Nosal, C. S. (2009). The structure and regulative function of the cognitive styles: a new theory. *Polish Psychological Bulletin, 41(3)*, 122-126.

Nosal, C. S. (2010). Ewolucja intuicji i jej funkcje w umyśle człowieka. In: A. Motycka (Red.), Życie na czas: Perspektywy badawcze postrzegania czasu (pp. 365-397). Warszawa: *Wydawnictwo Naukowe PWN.*

Nosal, C., Sobków, A. (2012). Wstępny raport z badań nad Kwestionariuszem Stylów Poznawczych. Wrocław: *SWPS*

Sexton, D., Bowman-Upton, N. (1991). Entrepreneurship, creativity, and growth. *Macmillan*, New York.

Sobków, A. (2014). Podstawowe mechanizmy zdolności intuicyjnych: ich struktura i relacje do otwartości umysłu, inteligencji oraz podatności na inklinacje poznawcze. Wrocław: *SWPS*

Spinnler, & W.A. Wagenaar (Eds.), *Theoretical perspectives on autobiographical memory* (pp.167– 193). Dordrecht: Kluwer Academic.

Sternberg, R., Davidson, J. (1996). The nature of insight. Massachussets Institute of Technology.

Sternberg R., J. (2004). Successful intelligence as a basis for entrepreneurship. *Journal of Business Venturing, 19, 189-201.*

Tubek, M., Piskorz, Z. (1994). Myślenie pojęciowe – pomiar i korelaty. *Przegląd Psychologiczny 187-194.*

Advancing ACTA: Developing Socio-Cognitive Competence/Insight

Julie GORE[a], Adrian BANKS[a], Almuth McDOWALL[b]

[a] University of Surrey, UK [b] Birkbeck College, University of London, UK

ABSTRACT

Accelerating the cognitive expertise of engineering professionals is a critical challenge for many high reliability, international organizations. This paper reports a collaborative, longitudinal, academic practitioner project which aimed to elicit, document and accelerate the cognitive expertise of engineering professional working with the manufacture and management of petroleum additives. 25 engineering experts were trained by three academic psychologists to use applied cognitive task analysis (ACTA) interview techniques in order to document the cognition of their expert peers. Results had high face validity for practitioners who elicited hot/sensory based cognition, a number of perceptual skills and mental models, highlighting undocumented context specific expertise. We conclude from a peer review of findings combined with experienced CTA analysts that ACTA techniques can be advanced in context by the explicit recognition and development of socio-cognitive competence /insight.

KEYWORDS

Applied cognitive task analysis, engineering, expertise, socio-cognitive competence/ insight.

INTRODUCTION

To date the naturalistic decision making (NDM) community have reported the strengths of applied cognitive task analysis (ACTA) and associated cognitive task analysis (CTA) techniques (Hoffman & Militello, 2008; Roth, 2008; Militello, Wong, Kirschenbaum & Patterson, 2011) which aim to capture and translate tacit cognition, developing new and important insights about how people are completing tasks. More recently these techniques have also begun to steadily grow in other research areas of organisational behaviour and management practice (Gore and McAndrew, 2009; McAndrew and Gore, 2012; Osland, 2010, 2013). Reports which focus upon the training of practitioners to adopt such methods and techniques however are less well documented. This work continues to examine the importance of the role of academics translating methodological research developments for impact and explorations *of* and *in* professional knowledge management practice (Anderson, 2007). We also note the importance of Vygotsky's (1978) assertion that the only way to understand how humans come to *know* is to study learning in an environment where *process* of learning rather than the product of learning that is the result of learning is studied. In addition, we aimed to ensure that aspects of cognitive expertise that are difficult to articulate were documented with clear application validity.

Organizational Context

A joint venture between ExxonMobil and Shell, the participants' workplace, is a leading organization in the formulation, manufacture and marketing of petroleum additives for lubricants and fuels. Shell has a long history of innovation in decision making and has effectively used scenario planning (Wack, 1985) for more than 45 years (see Wilkinson & Kupers, 2013 for a recent review). Shell's scenario practice began by exposing and questioning the future and facilitated dialogue in which managers' assumptions could safely be shared, questioned and challenged. Many business units and different organisational functions besides strategy and finance went on to develop scenarios which focussed upon the big-picture. In the 1980s however, a refocus was required which concentrated on '*deep listening*' in order to uncover uncertainties, probing the core concerns of leaders. Scenarios have continued to evolve and Shells' scenario developers aim to keep scenarios relevant and challenging learning tools which have impact upon organizational thinking and cognition.

Set within this innovative organisational culture the authors' were invited to explore within a much wider organisational project on knowledge management, *how best expert cognition in engineering expertise could be elicited, documented and shared, aiming to provide knowledge which would accelerate novice engineers' complex cognitive decision making processes.* Whilst Shells scenarios are most often at a macro-level of analysis this case organization was concerned with capturing expertise at the level of the individual. A key challenge here was to ensure the practitioners' accurately captured cognition in order to maintain continuous knowledge transfer within this highly qualified workforce. This paper documents the process of training transfer.

Expert cognition associated with managing uncertainty is highlighted (Lipshitz & Strauss, 1997) and aspects of hot/sensory based cognition explored. Notably, we offer suggestions for adapting and improving the CTA methods for management practitioners and highlight the importance of developing socio-cognitive competence. This latter area as yet, has been unexplored within the NDM or management community of researchers in depth and echoes Hoffman (2014) call for further explorations of the social aspects of CTA. We also note the importance of translating the findings from CTA for knowledge management, future scenario planning, management learning development and echo a cognitive constructionist approach.

Applied Cognitive Task Analysis: Unpacking expertise

Researchers have commented on the nature of expertise for several decades, significantly, Chi et al, (1988); Ericsson & Smith, (1991); Feltovich, Ford & Hoffman, (1994) within both laboratory-based examination and naturalistic investigation, exemplified by the Naturalistic Decision Making (NDM) framework. It is also important to note that this body of research has highlighted that experts learn in four key ways (Koehler & Harvey, 2004):

- engaging in deliberate practice, often setting goals and criteria for evaluation;
- compiling extensive experience banks;
- obtaining feedback that is accurate, and timely; and
- enriching their experiences by reflecting on their experience and lessons learnt from mistakes.

Several categories of knowledge related to expertise discriminate experts from other by describing what experts know and others, including novices, do not. Declarative and procedural knowledge (Anderson, 1983) are more apparent in experts. Put simply, experts know more domain and task related facts. In addition researchers within the NDM community suggest that: strong perceptual skills (Klein and Hoffman, 1983) are an essential component of expertise in many settings, as are mental models with depth; sensemaking of associations; the ability to run mental simulations; richer mental models enable experts to quickly spot anomalies and problems and also formulate information seeking tactics to manage uncertainty. Alongside the above components NDM research in the field suggests that experts metacognitive processes ensure that they take into account their own individual strengths and limitations. (For a recent discussion about how to recognise "good" CTA –see Roth et al 2014)

METHOD

Stage one: A pilot one day (7 hour) briefing about the use of ACTA techniques was provided (Gore, 2013) for a small group of professionals with different areas of engineering expertise. During a second day one of the authors trained 3 engineers to use a selection of the ACTA techniques (Militello and Hutton, 1998). Stage two: a 3 day longitudinal (twenty one hours) training event completed over 3 months was provided by the 3 authors/CTA instructors for 22 engineering professionals (5 female, 17 male). The professionals had a range of engineering expertise in management, manufacturing technology, finance, human resources, information technology, product development and operations management. Many of the participants were senior research scientists educated to doctoral level, all with 5-15 years of domain specific experience (classified here, as domain experts).

Procedure

First, the researchers' completed a ***task diagram*** and ***knowledge audit*** in order to illustrate the interview techniques associated with stage one and two of ACTA. This process was stopped and re-started in order for the engineers to ask questions and clarify the process. The first stage of ACTA the production of a *task diagram*, provides the interviewer with a broad overview of the task. This interview helps identify areas requiring complex cognitive skills which can be examined in depth in stage 2 of the process: *the knowledge audit*. In order to identify the type of tasks which were seen to be essential by the expert engineers, task diagrams were completed for key areas of engineering work which involved cognitive complexity. It is this type of work the organisation recognised was not currently documented meaningfully in training procedures. The professionals (experts) involved in the knowledge management project were mindful that areas of expert cognition which would be elicited via ACTA would result in more explicitly documented knowledge, which would be ultimately transferred to novice engineers for training purposes. The interviewee (practitioner engineer) begins by asking the interviewee (expert engineer) to break down a cognitive task related to their expert job role into 3 to 6 steps. These steps/stages are documented via a flip chart/ cognitive map which show 3-6 circles which relate to the task. The interviewer then asks which step/stage of the task is most cognitively challenging and why may novices find this difficult. This first stage can take up to 30 minutes to complete. The interviewer is encouraged to check on understanding with the expert to ensure that she or he agrees that the task diagram accurately provides a broad overview of the task. Together the interviewer and interviewee identify which element of the task is most cognitively complex and takes most thinking, judgment and decision making. This stage of the task is then explored and probed in great detail by completing stage two of ACTA, the Knowledge audit.

Second, the engineers practiced knowledge audit techniques with each other and documented their understanding of complex cognition. Again, a stop – start approach was adopted to facilitate the question technique and the documentation of knowledge elicited. *The knowledge audit* focuses upon a cognitive sub-task elicited from the task diagram and is well documented in the research literature in expert-novice differences (Crandall et al, 2006). A series of well-developed questions which are based on extensive research on expert thinking form the focus of the knowledge audit (Militello & Hutton, 1998). This stage of the ACTA is iterative and can take up to two hours to complete, eliciting lived stories and scenarios from the experts being interviewed. An optional third stage, *the simulation interview* assists the understanding of participants' cognition within the context of a challenging scenario developed from the knowledge audit. Simulations may be paper based or computer-based exercises which can then be a given to several domain experts to explore macro-cognitive complexity. This can be useful for developing training recommendations and is an area of ongoing work with the organisation.

Finally, a cognitive demands table was completed by the engineers, providing an analytical summary of data elicited. The cognitive demands table is a useful summary which provides an analysis of key aspects of expert cognition within the domain context and also clearly illustrates which aspects novices may find difficult. By documenting difficulties and capturing key cues and strategies for success, tacit knowledge is clearly illustrated. In addition to providing training in the ACTA techniques we also provided a briefing about theoretical issues in decision making and an exercise to facilitate active listening and questioning skills, as most of the participants had not previously had experience of research-based interviewing and had a genuine interest in the theoretical roots of the CTA methods. All participants had no prior experience of intensive research based interviewing and completed a questionnaire evaluation of their training experience. This questionnaire was developed aiming to evaluate cognitive, skill-based and affective learning outcomes (Kraiger, Ford & Salas, 1993), providing construct-orientated evidence of validity. A peer evaluation of the application validity of the cognitive demands tables and training scenarios produced from the interviews was also completed in collaboration with experienced analysts. Additionally, data was checked with other engineering experts to establish how far they agreed with the cognition elicited and most importantly how far they concurred that this tacit information was not currently available to novices.

RESULTS

The engineers found the process of interviewing and being interviewed using the ACTA techniques initially challenging. The ability of both the interviewer as facilitator of cognitive knowledge elicitation, and the interviewee, to take time to reflect in a thoughtful, reflexive , meaningful and organised way were key to the success of the interviews. The participants found the training involved a great deal of focus which meant lots of thinking/rest breaks were required. As a result of this the authors and engineers developed a series of tips, shown in table 2 in order to maximise the task diagram and knowledge–elicitation phase of ACTA, recognising the importance of socio-cognitive competence/insight. This series of tips greatly assisted participants and added to the language and positive social context for knowledge transfer. The tasks covered by the managers/engineers varied according to their organizational role and included everything from plant trial management; complex decisions surrounded choice of experiments for fuel testing; running a new project; improving supply security; to preparing to meet a new customer. Each of the engineers reported that the knowledge elicited, including key cues for improving situation awareness and scenario planning had rarely been documented in such a pragmatic way previously.

In addition to documenting task specific mental models, detailed perceptions of cues and strategies, an important feature which emerged to the surprise of the engineers was the importance of *hot/sensory based cognition*. For example several engineers described noticing peculiar smells in the mornings which resulted in adjusting the manufacturing process before the new petroleum additive was destroyed, making significant economic savings and avoiding potential hazards. The completed summary analysis/cognitive demands table were then used as a base for developing computer-based training which captured the lived expert realities of successful engineering tasks, clearly documenting mental models. The results of each ACTA were also subject to peer review which assisted knowledge transfer.

Feed-back from within the organisation has been positive with the practitioners wanting to utilise more CTA based training which provides such positive impact on organizational learning. The evaluation of the training suggested that the majority of participants felt that the ACTA techniques were *a very effective and efficient framework for helping articulate how experienced colleagues do specific tasks, provided structured learning and clear training outcomes*. One participant however, suggested that applying ACTA maybe particularly difficult in terms of *"drilling down to the right level of granularity of a task in order to access the most specific tacit knowledge".*

Table 2 Accessing Expertise in the Field: Top tips for getting rich data/ developing socio-cognitive competence/insight from ACTA

Redo and refine the task diagrams	Retrace your steps and redo the task diagrams as needed – you may need several drafts to get the detail level right
Listen actively throughout	ACTA works better if the interviewer listens actively: listen, summarise and then record the information (rather than writing notes throughout, as you are more likely to miss key information, particularly for the knowledge audit)
Stay focused and be clear about your roles	Reign in the temptation to share anecdotes, this can distract from the task, and remain clear about how interviews and who is interviewer (rather than inadvertently swopping during the process)
Bear with frustration	The process might entail some frustration about taking too long, or not getting the right level of detail– this is completely normal! If in doubt or getting too tired, leave the task for a while, and come back to it the next day
Ask what is difficult and ask about thinking	One of the key objectives for ACTA is to highlight what experts think, but might not have shared explicitly. So don't be shy to clarify, ask for more detail, or ask questions again in a different way. Your data should tap into thinking (so go beyond obvious outcomes)
Don't assume and choose your pairings wisely	You might think that things are obvious (as interviewer or interviewee) but chances are that they are not. It can work well to work in pairs or triads who don't usually work with each other, rather than pairing up with close colleagues. This will allow you to ask important questions which team members may not ask, assuming that the answers should be obvious (they usually are not!)
Remember that detail is good	As a rough rule of thumb, each component of your task diagram should be annotated with detail, and each aspect of the knowledge audit should fill about half a flip chart page
Be aware of when you stop recording information	If there is a time in the interview when you talk, but no information is recorded on the flip charts, then ask yourself 'why'. Are you not asking the right questions? Have you gone 'off track'?
Use the crib sheets	ACTA works best with structure, so don't be shy to use the crib sheets
Check your thinking	Do talk each other through your diagrams and knowledge audits again, for instance clarify anything which is not clear, and make sure the examples are specific, rather than general.

Limitations

Whilst ACTA and CTA techniques are established methods within the Human Factors, Psychology and Naturalistic Decision Making communities both with researchers and practitioners, few management researchers as yet, have adopted these techniques. Of the various perspectives which study judgment and decision making NDM has arguably made the greatest progress in industrial-organizational (I-O) psychology (Salas, Rosen and DiazGranados in press). Time intensive research activity is a 'nice to have' for many organisations and the management community may require these techniques to be adapted and modified further in order to translate to different domains of management practice. Evaluating the success of CTA based training requires a longitudinal approach which with this study we begin to offer. Continued research in this area also requires a shift in thinking and long term investment by more organisations in order to successfully manage knowledge learning transfer (Wang, 2010). In addition as the in-depth interview techniques are intensive and access System 2 thinking/cognition to reflect upon System 1 thought/cognition processes, careful interpretation, and mentoring is required. (Systems 1 thinking is characterized by fast, heuristic-based, emotional processing and is generally social and personal, and System 2 thinking is characterized by slower, controlled, analytical processing and is less social and less contextualised (Stanovich & West, 2000).

CONCLUSION

Tofel-Grehl and Feldon (2013) have noted the growing popularity of cognitive task analysis (CTA) in both research and practice and completed a meta-analysis of studies in order to examine the value of such training. They report that though their meta analysis is limited due to its small number of studies the effect of CTA instruction is large (Hedges's g=0.871). Also, whilst they note that effect sizes vary by CTA used and by training context our work to date concurs with their report and suggests that expert engineering information elicited with ACTA provides a strong basis for the highly effective training of novices. Whilst this work is ongoing it aims to be original in its application as few studies document such applied inclusion of practitioners with the co-construction of knowledge. The study demonstrates the utility of: applying qualitative methods such as ACTA to the domain of petroleum management/engineering; understanding how engineering practitioners' can adopt and utilise ACTA techniques; developing & interpreting the co-construction of knowledge management within a macro-cognitive framework . The elicited scenarios will aim to assist novice engineering professionals: raise situation awareness in relation to specific tasks and clearly define cognitive complexity in an organisational based repository of training scenarios.

Further, more detailed work is currently being completed in this area which should support knowledge management development (Donate & Canales, 2012) within the organisation. In addition, further work needs to be completed to assess if all of the professional engineers can easily utilise the ACTA techniques, assisting organisational learning in order to provide transformative innovations to knowledge management and support macro-cognitive awareness.

Our contribution to the development of CTA methods and knowledge management impact here strongly highlight the importance of recognising, managing and providing training which supports practitioners to develop their socio-cognitive competence and insight, alongside knowledge elicitation documentation and transformative knowledge management solutions. The complexities surrounding such knowledge transfer provide an interesting research agenda which utilises a range of theoretical and pragmatic contributions. Exploring the links theoretically between developing the reflexive System 2 thinking that the ACTA techniques require in order to reflect upon System 1 thinking also offer an exciting research agenda. We still have many more questions to answer however, concur with Hoffman et al (2014), and agree that (i) developing robust methods to accelerate expertise within organisations in order to assist knowledge acquisition and skills at a high level of proficiency in addition to (ii) facilitating the retention on knowledge and skill, will remain important to the future success of organisations training for resilience and adaptivity.

REFERENCES

Anderson, N. (2007). The practitioner-research divide re-visited: strategic-level bridges and the role of IWO psychologists. *Journal of Occupational and Organizational Psychology*. 80, 2,175-183.

Chi, M.T.H., Glaser, R., & Farr, M.J. (eds.) 1988. The Nature of Expertise. Mahwah, NJ: Lawrence Erlbaum

Associates. Crandall, B., Klein, G., & Hoffman, R. H. (2006). *Working minds*. Cambridge: MIT Press.

Donate, M. J., & Canales, J. I. (2012). A new approach to the concept of knowledge strategy. *Journal of Knowledge Management, 16* (1), 22-44. Ericsson., K.A. & Smith, J. (1991). *Towards a General Theory of Expertise: Prospects and Limits*. Cambridge,

MA: MIT Press Feltovich, P.J., Ford, K.M., and Hoffman, R.R. (eds.) (1997). *Expertise in Context*. Menlo Park MA: MIT Press Gore, J. (2013). Transforming Knowledge: Capturing Engineers' Cognitive Expertise. *11th International Conference on Naturalistic Decision Making*, Marseille.

Gore, J. & McAndrew, C. (2009). Accessing expert cognition, *The Psychologist*, 22, 3, 218-219. Hoffman, R.R., & Militello, L.G. (2008). *Perspectives on cognitive task analysis: Historical origins and modern communities of practice*. New York, NY: Taylor & Francis.

Hoffman, R.R.,Ward, P., Feltovich, P.J., Dibello, L., Fiore, S., & Andrews, D.H. (2014). *Accelerated expertise*. New York: Psychology Press, Taylor & Francis.

Kahneman, D., & Klein, G.A. (2009). Conditions for intuitive expertise: a failure to disagree. *American Psychologist.* 64(6). 515-526.

Klein, G., & Hoffman, R. (2008). Macrocognition, mental models, and cognitive task analysis methodology. In J. M. Schraagen, L. Militello, T. Ormerod, & R. Lipshitz, (Eds.). *Naturalistic decision making and macrocognition.* (pp. 57-80). Ashgate: Hampshire, U.K.

Koehler, D.J. and Harvey, N. (2004). Expertise in judgement and decision making : a case for training intuitive decision skills in *Blackwell Handbooks of Experimental Psychology : Blackwell Handbook of Judgement and Decision Making.* London, Blackwell.

Kraiger,K., Ford.J.K. & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology* Vol 78. No2, 311-228.

Lipshitz, R. and O. Strauss (1997). Coping with Uncertainty: A Naturalistic Decision-Making Analysis. *Organizational Behavior and Human Decision Processes,* 69, 2,149-163

McAndrew, C. & Gore, J. (2012). Understanding preferences in experience-based choice: a study of cognition in the wild. *Journal of Cognitive Engineering and Decision Making.*

McAndrew, C. & Gore, J. (2010). "Convince Me..." An Inter-Disciplinary Study of NDM and Portfolio Managers. In Mosier, K.L. & Fischer, U.M. (Eds.) *Informed by Knowledge: Expert Performance in Complex Situations.* Psychologist Press , 353-36.

APPLYING THE PRINCIPLES OF MAGIC AND THE CONCEPTS OF MACROCOGNITION TO COUNTER-DECEPTION IN CYBER OPERATIONS

Simon HENDERSON[a], Robert HOFFMAN[b], Larry BUNCH[b], Jeff BRADSHAW[b]
a. Centre for Cyber Security & Information Systems, Cranfield University, Defence Academy of the UK
b. Institute for Human and Machine Cognition, Florida

ABSTRACT
Like magic tricks, most cyber attacks involve some form of deception. What are the key factors in cyber deception and how can we characterize and anticipate them? Concepts of macrocognition and the theory of magic are formative of a scheme to support cyberworkers as they try to make sense of complexity and dynamics, and act effectively in the face of uncertainty. This paper outlines a general theoretical foundation and multifactorial analytical scheme for the analysis of cyber attacks. In our primary case study we analyze the 4chan hack of the Time Magazine "Person of the Year" web poll. To demonstrate the extensibility of the scheme, we also deconstruct password cracking, footprinting, key logging and buffer flow cyber attacks.

BACKGROUND

Like magic tricks, cyber attacks involve deception, even when deception is not the sole purpose of the cyber attack. The primary purpose or intent of cyber attacks is of course to achieve some effect, and that intent can be enabled by deception. Most of the salient examples of cyber attacks involve deception that is brought about by various means. Deception is defined as a deliberate action to induce erroneous sensemaking and subsequent activity within a target audience to achieve and exploit an advantage (Henderson, 2011). The purpose of the deception is to bring about some influence on the defender, either to get the defender to do something or to keep the defender from doing something (e.g., noticing the attack).

Numerous treatises have been written on deception in military campaigns (see Cruickshank, 1979; Dewar, 1989; Gooch & Perlmutter, 1982; Holt, 2008; Howard, 1992; Latimer, 2001; Whaley, 2007). There are also numerous and respected discourses on the art and theory of deception that is employed in magic (Earl, 2012, 2013; Fitzkee, 1943, 1944, 1945; Higham, 2009, 2011; Jermay, 2003; Lamont and Wiseman, 1999; Ortiz 1994, 2006; Steinmeyer, 2004; Tamariz, 1987; Triplett, 1900). There are scientific studies of the rules of persuasion and influence (Cialdini, 2001), and there are useful analyses of con artistry (Cornelius, 2009; Lovell, 1996; Maurer, 2000; Robbins, 2008). The principles and analytical scheme presented here were originally conceived for the context of military deception planning (Henderson, 2007, 2011, 2012). Here, these ideas are synthesized into an operational procedure for defense against deception specifically in cyber defense.

A cyber attack can happen on a temporal scale that is so brief that it precludes human comprehension, analysis and intervention. Thus, one might assume that the only possible solution for cyberdefense operations is solely computational in nature. However, all forms of deception basically involve humans trying to mislead other humans. As the magician and theorist Daniel Fitzkee (1943) said,

"Ultimately it is the spectator's mind which must be deceived, or there is no deception whatever. All of the apparatus we use, all of the secret gimmicks we employ, all of the sleights and stratagems we invoke—everything which identifies magic as mystery—the whole is designed to deceive the mind, and the mind alone, of the spectator."

Setting aside the ways in which cognitive work might play into cyber defense at the micro-temporal scale, cognitive work absolutely plays into cyber defense as macro scales where strategies and tactics are crucial.

DECEPTION CAN BE ANALYZED IN TERMS OF THE MAGICAL PLOYS

Discourses on magic have discussed dozens of magical "ploys." A ploy involves misleading the audience into thinking that their search for information is adequate and has been satisfied (Lamont and Wiseman, 1999). A clear example of a ploy is the "cover," when the magician waves one hand on a broad movement, thus distracting attention from what his other hand is doing. Additional ploys are the "glance" (people will look where the magician is looking), the "missing page" (tell a story that has obvious gaps), the "surprise" (violation of expectations) the "exploit" (tapping biases expectations or prior beliefs), and the "reveal" (to make something obvious). Ploys can be understood in terms of which macrocognitive functions or processes they manipulate or disrupt (e.g., sensemaking, attention management, storytelling, etc.).

DECEPTION LEVERAGES MACROCOGNITION

Deception is achieved through the presentation of cue sequences that, via pattern recognition (framing), influence the process of mental modeling and thereby influence decision making. An attacker is able only to stimulate or direct a target's attention and problem detection. However, by manipulating attention in a structured way, the defender's mental modeling can be influenced. For example, repeated cue sequences can be used to condition a target's expectancies, pattern development, and direction of attention. Furthermore, if expectancies derive from the defender's own mental model—that is, the defender has no awareness that their mental model has itself been influenced—then the defender's expectancies will be vulnerable to deception. Another principle is that d*eception is more success*ful if it includes some form of emotional state induction, which can induce time pressure and interfere with reasoning.

These, and additional principles apply not just to individuals but to targets that operate on the basis of collective belief, that is, team and organizational sensemaking. Furthermore, the principles of deception necessarily invoke the fundamental macrocognitive processes and functions (see Klein, 2007; Klein, Moon and Hoffman, 2006; Klein, et al., 2003), especially sensemaking, mental modeling, and projection to the future. Cyber work brings additional macrocognitive functions into the mix, especially problem detection, flexecution, management of uncertainty, and management of attention.

From the cyber defender's perspective, the primary challenge for attention management is to answer the question *"Where do I look to find the data I need?"* From the cyber attacker's perspective, the defender's attempt to answer this question can be influenced by directing the defender's attention, dividing the defender's attention, creating noise, exploiting the defender's inattention, concealing information, denying the defender the opportunity to find information, hiding the information, simulating the information in the "wrong" place, revealing false or bogus information, or substituting believable information for the genuine information. All of these are ploys utilized by magicians.

From the cyber defender's perspective, the primary challenge for sensemaking is to answer the question *"What counts as data?"* From the cyber attacker's perspective, the defender's sensemaking activity can be influenced by making a moving target, revealing the data and in so doing lead the defender to believe that the data must be deceptive, inducing confirmation bias on the part of the defender (the defender seeks information that confirms their hypothesis), inducing disconfirmation bias (the defender does not seek information that would disconfirm an hypothesis), or swapping reality for an obvious and bogus deception (the "double-bluff").

Once the defender has an initial frame that determines what counts as data, the question for sensemaking is "*How do I understand these data?*" Will the initial frame be questioned and refined or questioned and rejected? From the cyber attacker's perspective, the defender's sensemaking activity can be influenced by suggesting a pattern, supporting the verification of expectations, repeating pattern fragments to condition expectations, meeting the defender's expectations, dazzling (distracting) the defender, feeding the defender piecemeal information in order to stretch out the defender's sensemaking process, "accidentally" exposing the attacker's intent, such that the defender does not believe it, or fragmenting the pattern (make the defender invest effort to figure things out, thereby increasing the strength of their attachment to derived erroneous conclusions).

Once the defender has a frame that determines what counts as data, and the frame has been confirmed, or refined and improved, the question for is *"How do I act on these data?"* From the cyber attacker's perspective, the defender's flexecution activity can be influenced by falsely confirming the attacker's intent and thereby causing the defender to engage in the wrong actions, falsely confirming that the defense has been effective, thereby causing the defender to cease an action, constraining the effectiveness of a defense, delaying the defender's actions, or channeling the defender's actions in certain directions or certain kinds of activity.

CASE STUDY

'Emily Williams' was a 2012 internet-based social engineering and technical attack conducted by two security researchers (Lakhani and Muniz, 2013) to gain access to a US Government VPN, take control of their email system, obtain access to confidential information, and obtain a physical laptop belonging to the organization. The attack was based on 'Robin Sage', another fictitious person created in 2009 as a demonstration in the ease of obtaining information from intelligence on US military personnel via social networks; the successful Robin Sage findings were presented at Black Hat 2010 (Ryan, 2010).

The researchers first created a false Facebook and LinkedIn profile for a character they named 'Emily Williams' ('Control Attention' via 'Planting', 'Show the False', 'Fragment Story Fragments'). An attractive waitress (exploiting Cialdini's notion of 'Liking') volunteered photos for the fictitious character. She actually worked at an establishment frequented by the target company's employees (the nearby Hooters) yet no employee recognised her in person at any time during the experiment.

Before targeting the government target's employees, Lakhani and Muniz built Williams's presence on social media, building hundreds of connections ('Show the false' via 'Inventing', 'Social Proof'), with only one man flagging her as suspicious. Another man asked how Emily might know him, and when the researchers answered with information they obtained from the man's social media profile ('Anticipate Suspicion Driven Searching', 'Show the False' via 'Mimicry), he said he did indeed remember the imaginary girl ('Memory is Attention in the Past').

Once Williams had friends, the researchers updated her Facebook and LinkedIn profiles with just-hired status at the government target ('Mimicry', 'Generate Expectations'), and gave her an engineering title ('Authority'). The attractive, imaginary young woman connected with the target's employees via social media and connected with Human Resources, IT Support, Engineering and those in executive leadership roles (further 'Social Proof'). The congratulations for "her" new job subsequently rolled in.

As it was near the holidays, no one questioned when Williams posted seasonal cards to Facebook directed at specific targets among her co-workers - which they clicked, executed a Browser Exploitation Framework (BeEF) signed Java applet that opened a reverse shell back to Lakhani and Muniz via an SSL connection ('Liking', 'Reciprocity', 'Hide the Real' via 'Repackaging', 'Control Expectations', 'Simulation the Action').

Key logging was then used to gather passwords and insider information to gain access to the target agency ('Hide the Real' via 'Repackaging'). The researchers were able to figure out domain credentials to create an inside email address for Williams ('Show the False' via 'Invention'), VPN passwords to gain internal access and other methods to compromise the target.

The use of an inside email account subsequently enabled further social engineering ('Show the False' via 'Invention', 'Exploit Prior Beliefs'). Men working for the government agency were targeted to provide Williams, special treatment based on her attractive photograph ('Liking'). Some men offered to help Miss Williams at her new job by doing her a few favours; namely circumventing usual channels to get her a work laptop, and access to the organisation's network ('Pique Curiosity/Lure', 'Emotional Appeal').

We selected Operation Emily because it demonstrates how a single cyber attack episode can involve multiple strategies and multiple forms of deception. Thus, it serves as a useful case for the analysis based on the Theory of Magic. One should not assume the stereotype of the cyber attack as a single entity launching clever software and then just sitting back to see what happens. Cyber attacks are much more like army-on-army conflicts or like races. An attack can involve multiple and independent or even competitive hacker entities.

TOOL FOR ANALYSIS IN CYBER DEFENSE

The above ideas can be composed as a tool for cyber defense operations. The basic scheme is a matrix with columns such as:

- Attacker's goal/intent,
- Attacker's ploy,
- Attacker's actions to implement the ploy (what changes or moves?),

- Defender's indicators of an attack,
- Defender's counter indicators (that there is no attack),
- Defender's actions to prevent or mitigate the attack,
- Defender's indicators that prevention has been achieved,
- Attacker's indicators of attack progress,
- Attacker's indicators that a defense has been engaged,
- Attacker's bogus indicators of the success of a defense, and
- Defender's bogus indicators of a defense success.

The rows of the matrix would be specific cyber attacks. This is illustrated in Table 2, below. This Table is for illustrative purposes; it includes only four types of cyber attack and only the first four columns in the complete matrix.

Table 1. A scheme for counter-strategies in cyber defense.

1. Deceiver's Goal/Intent	2. Deceiver's Ploy	3. What changes or moves?	4. Defender's Indicators
ATTACK: Footprinting			
To acquire network, service and layout information; to map potential targets in the network.	Conceal (passive scanning); Camouflage (active scanning); Exploit inattention.	Connections are made from the attacker site.	Noise is created (Defender notices connection failures and hits into a darknet); Changes in the number of servers that are providing a service; Changes in the types of servers that are providing a service; Change in communications protocol that a server is using; Suspicious behavior on the part of a client (repeated failures).
Attack: Password Cracking			
Access privileged information.	Camouflage; Making a moving target.	Information is transferred off the target's network.	Periodic password login failures over a number of different users.
Attack: Key Logging			
Obtain personal information.	Camouflage (a process name that target does not recognize).	Accounts, passwords, financial information from the target computer back to the attacker	Virtually none; Target must know the attack is happening and look for it.
Attack: Buffer Overflow			
Take control of target's computer.	Exploit inattention.	Information packets are sent from the attacker's computer to the target's host computer; Overwrite of existing target information.	Bad packet.

GENERALIZED MODEL

This analysis suggests a general event model, depicted in Figure 1. This diagram covers only selected aspects of what we call the "Three Cycles." In Cycle One, an attack is launched, it is detected or not, and it succeeds or not. In Cycle Two, there is defense activity and deception activity, either of which might or might not be successful. In Cycle Three there is counter-deception and the use of bogus ploys and bogus indicators. We refer to these as Cycles because they involve closed loops, that is, they all have feedback implications (e.g., a defensive operation might be observable by the attacker and hence "give things away"). Obviously, Cycle Three is where things get highly complex and confusing (see Hoffman, et al., 2011).

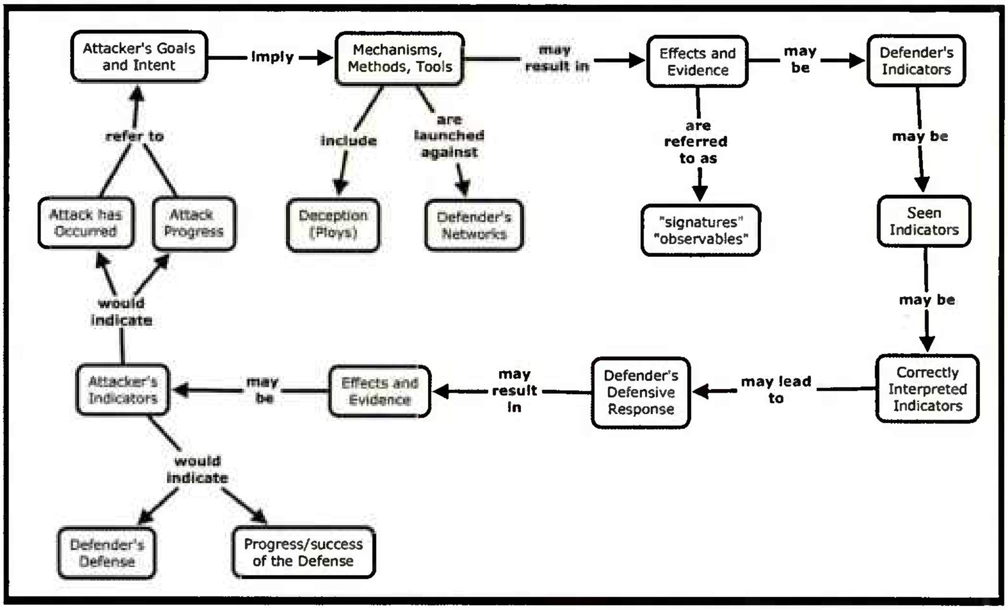


Figure 1. A process description of for cyber defense and offense based on the theory of magic and the concepts of macrocognition.

At a recent DaD meeting, a senior intelligence officer was asked:

Shauld intelligence aperotions olwoys ossume thot the ottocker is being deceptive, ond that the ottacker knows thot the defender believes thot the ottocker is being deceptive, ond that the defender will engoge in counter-deception, ond thot the defender should engoge in counter-deceptian? And can't this oll not drive you nuts, butyau have ta think it through this deeply?

The officer's answer was "Yes." Our current effort invalves applying this analytical scheme to additianal cyber attack and defense activities, and fleshing out the descriptive models of the "Three Cycles."

ACKOWLEDGEMENT

Preparation of a draft af this paper at IHMC was suppOrted in part by a Subcontract to SoarTech, Inc., an an SBIR from the Office of Naval Research.

REFERENCES

Cialdini, R.B. (2001). The science af persuasian. *Scientific Americon, 284*, 76-81.
Cornelius, G. (2009). *The art af the can.* Alexandria, VA: American Correctional Assaciatian.
Cruickshank, C. G. (1979). *Deception in Warld War II.* New Yark: Oxford University Press.
Dewar, M. (1989). *The ort of deception in worfore.* Newton Abbot: David & Charles.
Earl, B. (2012, 2013). *Less is more* (Vols 1, 2) Benjamin Earl
Fitzkee, D. (1943). *Showmonship for Mogicions.* San Rafael, CA: San Rafael House
Fitzkee, D. (1944). *The trick broin.* San Rafael, CA: San Rafael House
Fitzkee, D. (1945). *Mogic by misdirection.* San Rafael, CA: San Rafael House
Gooch, J., and Perlmutter, A. (Eds.). (1982). *Militory deception ond strotegic surprise.* New York: Frank Cass and Company Limited.

Henderson, S. M. (2007). "Deception: A Guide to Exploiting the Psychological Basis of Deception in Military Planning." MIST/06/07/702/21/1.1. Farnborough: QinetiQ.

Henderson, S. M. (2011). Deceptive Thinking Workshop. Paper presented at the 1st MilDec Military Deception Symposium, 2nd-3rd November 2011, Defence Academy of the United Kingdom, Shrivenham.

Henderson, S. (2012). Military Deception: Learning From Other Domains. Presentation at the 2nd MILDEC Deception Symposium, Defence Academy of the United Kingdom, Shrivenham, 7-8 November 2012

Higham, J. (2009). *Secrets of improvisotionol mogic.* London: Justin Higham.

Higham, J. (2011). *The Kosbe system: The mechanics of improvisotion in cord mogic.* London: Justin Higham

Hoffman, R.R., Henderson, S., Moon, B., Moore, D.T., & Litman, J.A. (2011). Reasoning difficulty in analytical activity. *Theoretical Issues in Ergonomic Science, 12,* 225–240.

Holt, T. (2008). *The deceivers: Allied militory deception in the Second World Wor* (Vols 1, 2). London: The Folio Society Ltd.

Howard, M. (1992). *Strotegic deception in the Second World War.* London: Pimlico.

Jermay, L. (2003). *Building blocks.* Ashford: Alakazam Magic

Klein, G. (2007). Flexecution, Part 2: Understanding and supporting flexible execution: *IEEE Intelligent Systems, 22,* 108-112.

Klein, G., Moon, B. & Hoffman, R. R. (2006, November/December). *Moking sense of sensemoking 2: A mocrocognitive model.* IEEE Intelligent Systems, pp. 88-92.

Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (May/June, 2003). Macrocognition. *IEEE: Intelligent Systems,* pp. 81-85.

Lakhani, A., & Muniz, J. (2013). Social media deception. Paper presented at the RSAConference Europe 2013, October 29-31, 2013, Amsterdam.

Latimer, J. (2001). *Deception in wor.* London: John Murray.

Lamont, P. & Wiseman, R. (1999). *Mogic in Theory: An introduction to the theoreticol ond psychologicol elements of conjuring.* Hatfield, UK: University of Herfordshire Press.

Lovell, S. (1996). *How to cheot at everything: A con mon reveols the secrets of the esoteric trode of cheoting, scoms, ond hustles.* Philadelphia PA: Running Press Book Publishers

Maurer, D. W. (2000). *The big con: The story of the confidence mon ond the confidence trick.* London: Arrow Books

Ortiz, D. (1994). *Strong magic.* Washington, DC: Kaufman and Co.

Ortiz, D. (2006). *Designing mirocles* (Vol 1). El Dorado Hills, CA: A1 MagicalMedia.

Robbins, T. (2008). The Modern Con Man: How to Get Something for Nothing. New York: Bloomsbury

Ryan, T. (2010). Getting in bed with Robin Sage. [Retrieved 12/01/2015, from http://www.omachonuogali.com/BlackHat-USA-2010-Ryan-Getting-In-Bed-With-Robin-Sage v1.0.pdf]

Steinmeyer, J. (2003). *Hiding the elephont: How mogicions invented the impossible ond leorned to disoppeor.* New York: New York: Carroll & Graf Publishers.

Tamariz, J. (1987). *The magic woy.* Madrid: Editorial Frakson.

Triplett, N. (1900). The psychology of conjuring deceptions. *The Americon Journol of Psychology, 11,* 439-510

Whaley, B. (2007). *Strotegem: Deception ond surprise in wor.* Norwood, MA: Artech House.

TRUST AND RELIANCE AS EMERGENT PHENOMENA IN MACROCOGNITIVE WORK: AN INTEGRATED MODEL

Robert HOFFMAN[a], Bradley BEST[b], and Gary KLEIN[c]
[a] *Institute for Human and Machine Cognition*
[b] *Adaptive Cognitive Systems, Inc.*
[c] *Macrocognition, LLC*

ABSTRACT
Trust is an emergent from the dual processes of sensemaking the observed or controlled world and sensemaking the technology. Reliance is as an emergent from the dual processes of flexible execution of the macrocognitive work and flexible execution of the interactions with the technology. We present an integrated model that would lend psychological fidelity to computational models that might attempt to capture the richness of expert sensemaking and re-planning, and at the same time capture the richness and dynamics of trust in and reliance upon the automation that mediates the macrocognitive work.

INTRODUCTION
Macrocognitive models describe how cognition adapts to complexity (Klein, et al., 2003). Trust models describe how people develop trust in automation. These classes of models are based on empirical evidence about different phenomena, but those phenomena emerge in the same context: macrocognitive work. Our goal is to integrate these classes of models.

MODELS OF TRUST IN AUTOMATION
Models of trust in automation capture variables that influence trust in and reliance on automation (e.g., Oleson, et al., 2011). Some "models" are really lists of factors (e.g., Muir, 1987). Some models are boxes-and-arrows diagrams that express hypothetical causal sequences linking variables that determine an operator's state of trust in the automation (e.g., Lee and See, 2004). The goal of the models, especially the mathematical ones, is to estimate values of trust, treating it as a state variable, typically on the assumption of a fixed task. Is this all that we need a model of trust to do for us?

Trust in is complex and dynamic. Although there can be periods of relative stability, neither trusting (as a relation), trustworthiness (as an attribution), nor reliance (as an activity) is static. Relations develop and mature; they can strengthen slowly or decay rapidly. Furthermore, there are many varieties of trust relations, including mistrust and distrust (Hoffman, Johnson and Bradshaw, 2013; Merritt and Ilgen, 2008). *The trusting relation of humons to outomation involves some dynomic mixture of context-linked justified ond unjustified trust, ond context-linked justified ond unjustified distrust* (Hoffman, et al., 2009; Lyons, et al., 2011; Sheridan, 1980; Woods, Roth and Bennett, 1987). Thus, we regard the concept of trust as an entry point for the analysis of the usability, usefulness, understandability, and observability of work systems.

MODELS OF MACROCOGNITIVE WORK
The macrocognitive functions on which we focus are sensemaking and flexible execution, described by the Data/Frame model of sensemaking (Klein et al., 2006) and the Flexecution model of re-planning (Klein, 2007), presented in the two component panels of Figure 1 (following the References.) The boxes indicate macrocognitive functions and the arrows indicate dependence relations. The Data/Frame model describes what happens as people try to understand complex situations and continually work to refine and improve upon that understanding. The Flexecution model describes what happens as people try to achieve their goals even as they have to change their goals. Both models are composed entirely of closed loops. Causal chain models, such as Karl Duncker's classic model of hypothesis testing (Duncker, 1945) can be pulled out, but a major premise of the macrocognitive approach is that the functions of macrocognition are parallel, continuous, and interacting. There is good empirical evidence that the sensemaking and flexecution models must be merged. For example, in a study of the decision making by law enforcement officers, Ward, et al. (2011) showed that the re-consideration of a mental model does not cease even though a particular course of action has been adopted. For the analysis of cognitive work that is mediated by computational technology, there has to be even more: sensemaking and flexecuting the technology itself.

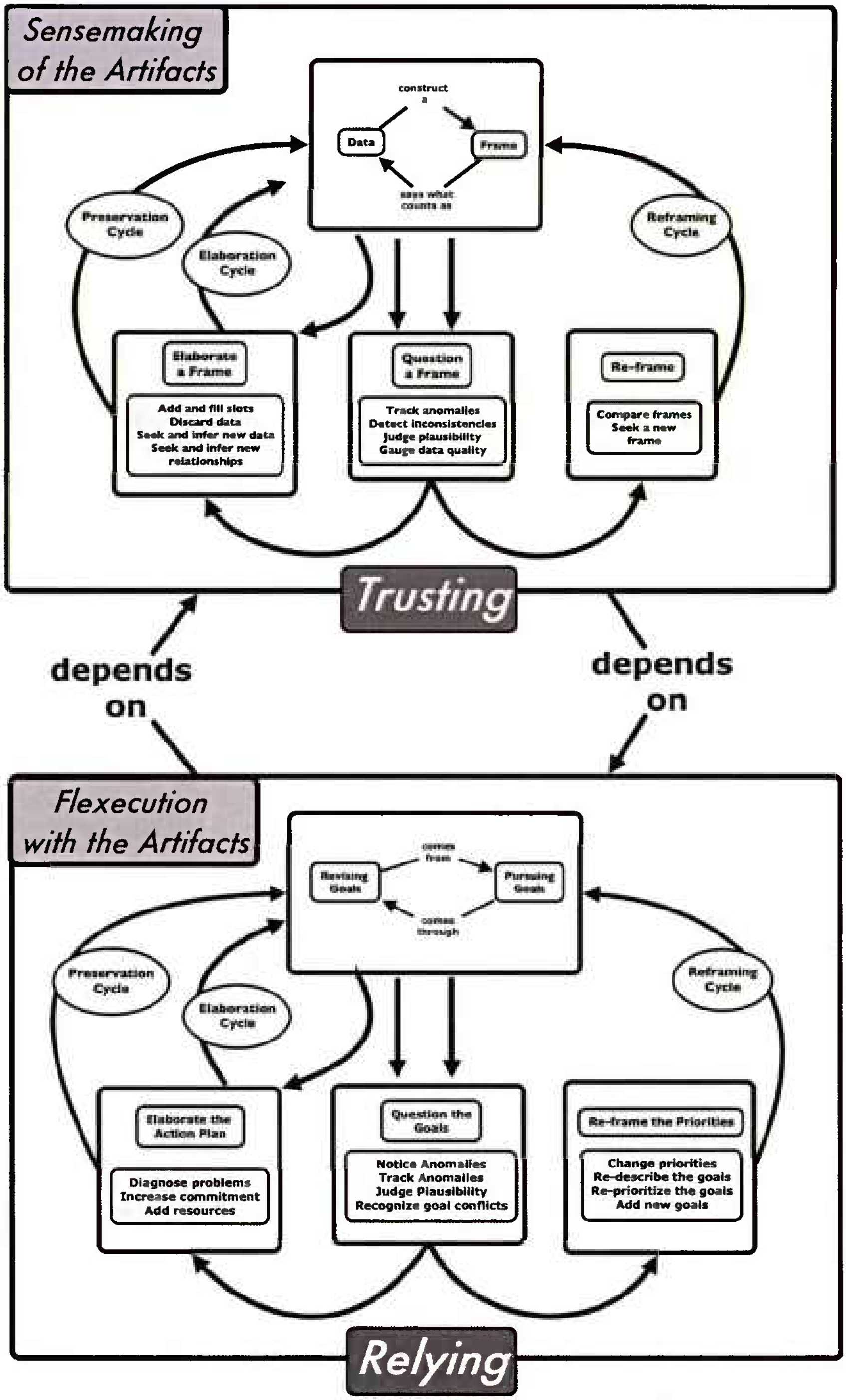


Figure 1. The dual Data/Frame Model of sensemaking and the Flexecution Model of re-planning, with regard to the Automation.

WORKING THE TECHNOLOGY

When conducting macrocognitive work, people try to make sense of the technology at the same time they are trying to make sense of the observed or controlled world. They have to learn what the technology does and why (Seong and Bisantz, 2002). People also have to flexecute in their interactions with the technology. They search for functionalities; they try to cope with the software's awkwardness by creating work-arounds and kluges (Koopman and Hoffman, 2003). A number of recent studies have shown how operators engage in sensemaking the technology and flexecuting their interactions with the technology as they judge the cognitive demands of tasks, and from that anticipate what the technology might do (e.g., Mosier, et al., 2012). The combination of Data/Frame and Flexecution modules (D/F+F) presented in Figure 1 is an attempt to express the dynamics of trusting in and relying upon automation. Basically, this loops two modules together by asserting their parallel interdependence.

THE INTEGRATED MODEL

As should be apparent, we need to double-up the modules if we are to have a complete theory of macrocognitive work. A high-level view of this quadruple is presented in Figure 2. *Issues of trust reside in the loop of sensemaking the world and sensemaking the technology, whereas issues of reliance reside in the loop of flexecuting the work and flexecuting the technology.* As is the case for the individual modules of sensemaking and flexecution, here we see nothing but closed loops. Sensemaking of the observed/controlled world depends on sensemaking of the technology. Flexecuting one's actions on the observed/controlled world depend on the ability to flexecute the technology. When each of the four nodes in Figure 2 is replaced by the corresponding detailed diagram (as in Figure 1), it becomes clear how this is a case study in complex systems (Feltovich, Hoffman and Woods, 2004; Walker, et al., 2010).

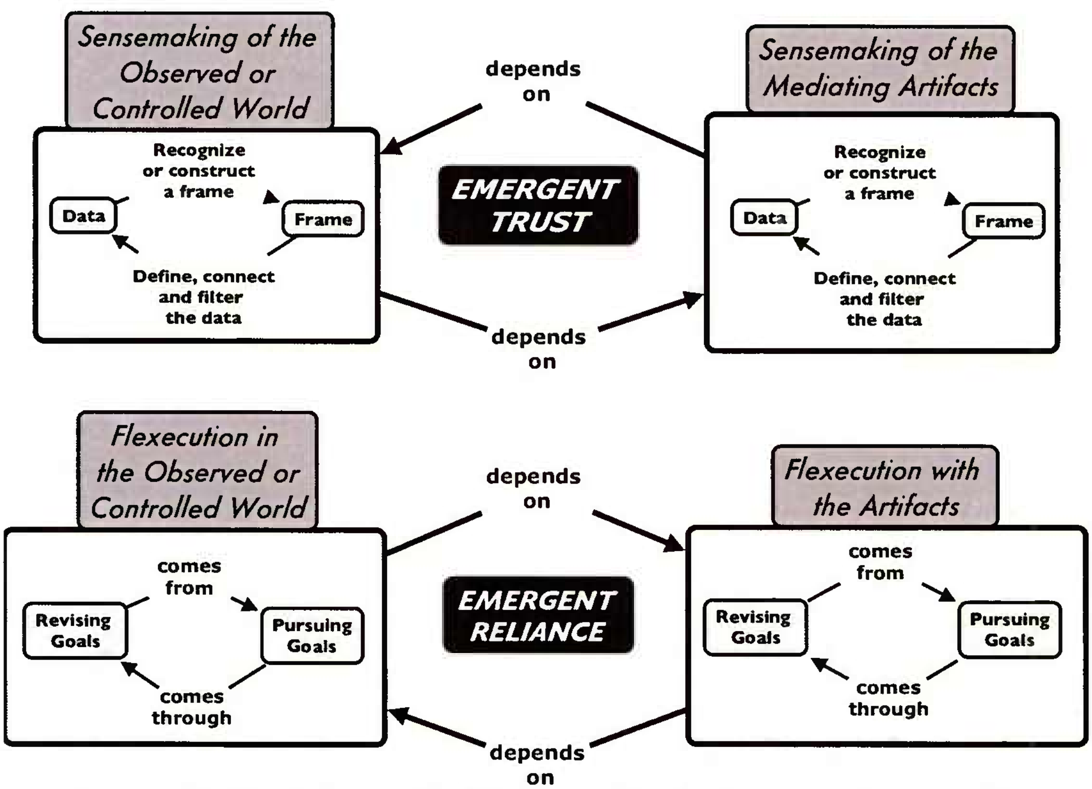


Figure 2. A high-level diagram of the D/F+F model. The details that are within each of the modules are only summarized here. The two right-hand modules in this Figure correspond

to the two modules in Figure 1, above.

Many factors have a causal or moderating influence on trust (prior knowledge, reputation, beliefs, gossip, etc.). Trust is an attitude, an attribution, an intention, a virtue, and an expectation. Trust is multidimensional and conditional; for instance, it can break down rapidly when the trustee (or machine) makes a mistake. A number of such relativities are generally recognized with regard to trust in automation (e.g., Lee and See, 2004). Person and knowledge variables influence temporal specificity (trust is conditional), functional specificity (trust is contingent), calibration (balance of justified and unjustified trust) and reliance. For example, in the model of trust in automation by Dzindolet, et al. (2002: see also Bisantz and Seong, 2001), person factors (self-confidence) and knowledge factors (past experience) lead to calibration (i.e., perceived risk, anticipated effort) and reliance subsequently depends on task factors (e.g., workload, time constraints). Recognizing those dependencies, the D/F+F model regards trusting as an emergent phenomenon, related to the operator's understanding of the technology in the immediate context of the work. Person and knowledge variables have to do with sensemaking (understanding the automation) and flexecution (learning how to work using and through the technology).

There isn't any node in the D/F+F model that says, "trust is computed here." The purpose of the model is not to calculate a trust state with a yes/no reliance outcome as the output of a causal chain. The purpose of the integrated model is to describe the macrocognitive work in such a way as show how trusting and relying can emerge and change in the ebbs and flows of deliberative attention to aspects of the cognitive work, and the feedback from the automation and the world that informs the worker of the effects of their actions. In a sense, trusting and relying are *everywhere* in the D/F+F model. The methodological implication is that the D/F+F model can do more than just output values of trust as a final state variable. One can insert judgment tasks or capture performance measurements at any time during a performance, and with reference to in any of the cycles described by the model, in an attempt to gauge momentary trust, calibration, reliance, etc.

In this way, the D/F+F model is perhaps complementary to other models. For example, the model is consistent with Muir and Moray's (1989) theory of machine trust, but is more comprehensive. On the other hand, the D/F+F model is orthogonal to the Parasuraman, Sheridan and Wickens' (2000) model of automation mistrust and misuse. As a microcognitive model, theirs proposes stages of information processing. Like Lee (2012), we question the generalizability of models that "consider the development and influence of trust in terms of a traditional information processing perspective where perceptual inputs are processed to affect trust and guide the decision to rely or comply with the automation" (p. 304). Our conjecture is that issues of trust are inherently wedded to the concepts of frame, context, and dynamics that are central to macrocognition. Information processing microcognitive models will be less effective in aiding our understanding of the phenomena.

PUTTING THE MODEL TO WORK

Methods of cognitive task analysis (CTA) trace reasoning processes and generate task or goal decompositions (see Crandall, Klein and Hoffman, 2006; Shepherd, 2001). None of the CTA methods explicitly and deliberately unifies the analysis of macrocognitive work, the analysis of trust in automation, and the analysis of reliance on the automation. The D/F+F model thus has a definite use: It enables us to chart paths, tracing the worker's reasoning as it flits from activity to activity within the macrocognitive work (to adapt William James' metaphor; James, 1890, p. 243). The activities described in the model can be taken as categories for coding a protocol. This leads to some interesting predictions. For example, it might be assumed that if attention has to shift away from the primary task goals and has to focus instead on making sense of the technology, that the cognitive work would suffer due to distraction or increased mental workload. Our presentation will highlight a case study of the reasoning of an expert weather forecaster, which suggests that this assumption may sometimes be incorrect: The forecaster's awareness of the capabilities and limitations of the technology, and methods for coping with such things as limited or sparse data, mesh seamlessly and effortlessly with his sensemaking of the weather itself. This pattern merits further empirical study, as does the search for other patterns entailed by the D/F+F model as a task-analytic scheme.

References

Bisantz, A. M. & Seong, Y. (2001). Assessment of operator trust in and utilization of automated decision-aids under different framing conditions. *Internotionol Journol of Industrial Ergonomics, 28*, 85-97.

Crandall, B., Klein, G. & Hoffman R. (2006). *Working Minds: A Practitioner's Guide to Cognitive Task Analysis.* Cambridge, MA: MIT Press.

Dzindolet, M., Pierce, L., Beck, H. & Dawe, L. (2002). The perceived utility of human and automated aids in a visual detection task. *Humon Factors, 44,* 230-244.

Duncker, K. (1945). On problem solving. *Psychologicol Monogrophs, 58*, 1-113.

Feltovich, P., Hoffman, R. & Woods, D. (May/June 2004). Keeping it too simple: How the reductive tendency affects cognitive engineering. *IEEE Intelligent Systems*, pp. 90-95.

Hoffman, R.R., Johnson, M., Bradshaw, J.M. & Underbrink, A. (2013). Trust in automation. *IEEE: Intelligent Systems*, pp. 84-88.

Hoffman, R.R., Lee, J.D., Woods, D.D., Shadbolt, N., Miller, J. &Bradshaw, J.M. (2009, November/December). The dynamics of trust in cyberdomains. *IEEE Intelligent Systems*, pp. 5-11.

James, W. (1890). *Principles of psychology.* New York: Holt.

Klein, G. (2007). Flexecution as a paradigm for replanning, Part 1: *IEEE Intelligent Systems, 22*, 79-83.

Klein, G., Moon, B. &Hoffman, R. R. (2006). Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems*, pp. 88-92.

Klein, G., Ross, K., Moon, B., Klein, D., Hoffman, R. & Hollnagel, E. (May/June, 2003). Macrocognition. *IEEE: Intelligent Systems*, pp. 81-85.

Koopman, P. & Hoffman, R.R., (2003). Work-arounds, make-work, and kludges. *IEEE: Intelligent Systems*, pp. 70- 75.

Lee, J. (2012). Rethinking *Trust in Automotion* (2004). *Proceedings Of The Humon Foctors ond Ergonomics Society 56th Annuol Meeting* (pp. 303-307). Santa Monica, CA: Human Factors and Ergonomics Society.

Lee, J. & See, K. (2004). Trust in automation: Designing for appropriate reliance. *Humon Foctors, 46,* 50–80.

Lewicki, R., McAlister, D. & 8ies, R. (1998). Trust and distrust: New Relationships and realities. *Acodemy of Monogement Review, 23*, 438-458.

Lyons, J., Stokes, C., Eschelman, K., Alarcon, G. & Barelka, A. (2011). Trustworthiness and IT suspicion: An evaluation of the nomological network. *Humon Foctors, 53,* 219-229.

Merritt, S. & Ilgen, D. (2008). Not all trust is created equal: Dispositional and history-based trust in human automation interactions. *Humon Foctors, 50,* 194-201.

Mosier, K., Fischer, U. & The HART Group (2012). Impact of automation, task and context features on pilots' perception of human- automation interaction. In *Proceedings of the Humon foctors ond Ergonomics Society 56th Annuol Meeting* (pp. 70-75). Santa Monica, CA: Human Factors and Ergonomics Society.

Muir, B. (1987). Trust between humans and machines, and the design of decision aids. *Internotionol Journol of Mon-Mochine Studies, 27,* 527-539.

Muir, B. (1994). Trust in automation Part 1: Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics, 37,* 1905-1922.

Muir, B. & Moray, N. (1996). Trust in automation, Part II: Experimental studies of trust and human intervention in a process control simulation. *Ergonomics, 39*, 429-460.

Oleson, K., Billings, D., Chen, J. & Hancock, P. (2011). Antecedents of trust in human-robot collaborations. In *Proceedings of the IEEE Internotionol Multi-Disciplinory Conference on Cognitive Methods in Situotion Aworeness ond Decision Support* (pp. 175-178). New York: IEEE.

Parasuraman, R., Sheridan, T. & Wickens, C. (2000). A model for types and levels of human interaction with automation. *IEEE Tronsoctions On Systems, Mon, And Cybernetics—Port A: Systems And Humons, 30,* 286-297.

Seong, Y. & 8isantz, A. (2002). Judgment and trust in conjunction with automated aids. In *Proceedings of the Humon Foctors ond Ergonomics Society 46th Annuol Meeting* (pp. 423-428). Santa Monica, CA: Human Factors and Ergonomics Society.

Shepherd, A. (2001). *Hierorchicol tosk onolysis.* London: Taylor & Francis.

Sheridan, T. (1980, October). Computer control and human alienation. *Technology Review,* pp. 61-73.

Walker, G., Stanton, N., Salmon, P., Jenkins, D. & Rafferty, L. (2010). Translating concepts of complexity to the field of ergonomics. *Ergonomics, 53*, 1175-1186.

Ward, P., Suss, J., Eccles, D., Williams, A. & Harris, K. (2011). Skill-based differences in option generation on a complex task: A verbal protocol analysis. *Cognitive Processes, 12*, 289-300.

Woods, D., Roth, E. & Bennett, K. (1990). Explorations in joint human-machine cognitive systems. In W. Zachary and S. Robinson (Eds.), Cognition, computing, and cooperation (pp.123-158). Norwood NJ: Ablex.

Tools for Facilitating Tacit Knowledge Elicitation using Critical Decision Method

Danny Shu Ming KOH[a], Hock Guan TEE[a], Boon Kee SOH[a] and Angela Li Sin TAN[a]
[a]*DSO National Laboratories, Singapore*

ABSTRACT

Critical Decision Method (CDM) is an interview technique commonly used for eliciting tacit knowledge in challenging and atypical complex situations. However, the knowledge elicitation process can be daunting as the interview process is more than often non-linear and opportunistic. We present a visualization tool configured from Microsoft PowerPoint that can be used to easily create an event timeline of the incidents described by the experts. We discuss how the features of this tool together with Cognitive Demands Table (CDM) spreadsheet can facilitate the interviewing process to ensure an accurate and comprehensive knowledge elicitation in the context of wargaming.

KEYWORDS

Cognitive Field Research and Cognitive Task Analysis; Military

INTRODUCTION

In order to understand how to perform the tasks effectively in specific professional fields, we rely on the respective Subject Matter Experts (SMEs) to provide us with the information. However studies have shown that self-reports from experts are often inaccurate and incomplete (Blessing & Anderson, 1996). The inability to articulate the procedures clearly is due to automaticity of knowledge (Feldon, Timmerman, Stowe & Showman, 2010). As the SMEs acquire expertise in a specific skill, less conscious monitoring is required in performing the skills, allowing him to perform the task quickly. While experts can verbalize the actions that they perform, the tacit part for them is to articulate the conditions of when these actions should be done.

Cognitive Task Analysis (CTA) is often used to ensure a comprehensive elicitation of the expert's knowledge. CTA is a general term that describes an inventory of techniques to elicit the related knowledge, cognitive processes and goal structures of performing a specific task (Chipman, Schraagen & Shalin, 2000). Interview techniques with SMEs, particularly Critical Decision Method (CDM) (Klein, Calderwood & Macgregor, 1989) are most commonly adopted by CTA practitioners (Tofel-Grehl & Feldon, 2013; Cooke, 1999). The insights from CTA interviews are often used to develop training materials (Feldon et al., 2011; Feldon et al., 2010), decision aiding tools (Hoffman, Coffey, Ford & Carnot, 2001) and also to inform the improvement of work processes and system design.

The interviewing process is however laborious and complex. CTA practitioner has to constantly identify critical decision points from the SME's responses and probe appropriate questions to elicit useful information without influencing SME's perception. Decision points are usually opportunistic and the CTA practitioner might have to halt the current set of questions and delve into the identified decision point. Upon eliciting the required information from the SME, the CTA practitioner would have to return to the previous set of questions to ensure the completeness of the information captured. The iterative knowledge elicitation process makes it difficult for the CTA practitioner to recall all questions that he needs to ask. Without a proper knowledge documentation tool, the CTA practitioner might be overwhelmed with the massive and unstructured set of knowledge provided by the SME. Failure to acquire a complete and accurate set of information during the CTA interview might consequently impede the development of effective systems or training materials. In view of the difficulties in eliciting knowledge from the SMEs through interviews, there is a need to identify or develop knowledge elicitation tools to facilitate these interviews. In this study we explored the feasibility of a visualization tool together with a recommended template for knowledge documentation while conducting a series of CDM interviews.

CRITICAL DECISION METHOD

CDM is an incident-based interview where the SME is asked to recall a highly challenging and unusual event that he has experienced. In essence there are four broad phases in the interview (Crandall, Klein & Hoffman, 2006). SME is required to share a personal incident that he felt was challenging at that time. If the incident is deemed appropriate and aligns with the objective of the CTA interview, the next phase is to elaborate the incident further by creating a timeline of events. The third phase is to probe into the critical decision points to identify the perceptual cues and alternative options when making a decision. These dimensions are aligned with what experts typically find difficult to articulate during self-reports. If time permits, the CTA practitioner will ask hypothetical "what-if" questions to understand how the decisions might change with varying conditions or situations. However the interview process is highly iterative and CTA practitioners are often cognitively challenged to ensure that the incident is appropriate and rich in content to elicit valuable insights.

COGNITIVE DEMANDS TABLE

As part of the recommendation by the Applied Cognitive Task Analysis framework by Militello and Hutton (1998), Cognitive Demands Table (CDT) is used to document elicited information from the CTA interviews. Given that not all information provided during the interviews will be important, the format provided by the CDT helps the CTA practitioners to quickly filter out irrelevant data. The CDT provides a standard set of headers that CTA practitioners can use to categorize the information gathered from their interviews. The headers are namely "Difficult Cognitive Elements", "Why difficult", "Common errors", and "Cues and strategies used". However these headers are not fixed and CTA practitioners are recommended to alter according to the information they would need in order to translate the information into development effectively.

While CDT proves to be an effective tool for knowledge representation during the interviews as well as for analysis phase, the knowledge represented is usually in the form of text. Text-based representation makes it difficult for users to comprehend the story and insights immediately. For instance, CTA practitioners might have to spend some time reading the text during the interview to identify a specific issue before they can show the information to the SME for discussion. Some form of the visualization of the events would potentially improve the usability of CDT, and subsequently enhance the work flow of CTA practitioners.

TOOLS TO FACILITATE KNOWLEDGE ELICITATION

There are several software tools developed to capture SME's knowledge. Radtke and Frey (1997) describes a procedure named "Sea Stories" that guides SME to "translate their conceptual knowledge and expertise into a representation" on a series of computer-based storyboards. The setback for this tool is that SME needs to learn a set of procedures in order to use the tool effectively. The SME will also have full control on what and how to translate his/ her knowledge into an appropriate form. However SMEs might not be able to provide a complete description of the knowledge due to automaticity (Blessing & Anderson, 1996) without an external intervention to probe deeper into the tacit knowledge.

VISUALIZATION TOOL

The challenges faced by the CTA practitioners serve as the impetus to develop a tool to aid CTA practitioners. We designed a scenario drawing and visualization prototype tool configured from Microsoft PowerPoint. The scenario drawing tool allows the CTA practitioner to quickly create snapshots of the critical events in the incident as described by the SME during the interview. By creating the images of the events, the CTA practitioner can better appreciate the incident as compared to an abstract verbal description by the SME. After the images of the events have been created, the tool can then compile all images and create an event timeline of the incident shared by the SME. The tool was designed for a series of CTA interviews in the context of military wargaming. However most of the features are context-free and can be used in other domains as well. The tool is and some of the key features are as follows:

Readily available maps and symbols

There is a set of maps and entity symbols that have been designed and installed in the tool. These maps and symbols are designed specifically to the context of wargaming. During the interview, the CTA practitioner can immediately choose the map relevant to the incident and create a replication of the critical events by placing the entities as described by the SME. The CTA practitioner can also create and save new entity symbols during the interview.

Textbox to allow detailed description of events

Textbox is also provided in the tool so that the CTA practitioner can make comments or observations that are difficult to represent in the scenario drawing tool.

Automated generation of event timeline

The tool allows the CTA practitioner to document the time of critical events mentioned by the SME during the interview. With the time provided for each event, the tool can immediately generate an event timeline of the incident in a slide. The timeline can then be presented to SME for discussion during the interview. The benefit of using the tool to create the event timeline is that the CTA practitioner can make any refinement or changes immediately without messing up the event flow. The scenario drawing tool and visualization of the event timeline is shown in Figure1 and Figure 2 respectively.

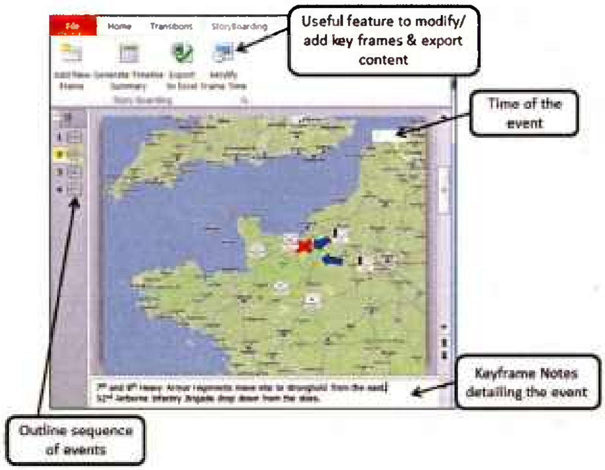


Figure 1. Features of the scenario drawing tool. The readily available maps, entity symbols and textbox allow CTA practitioner to quickly create the events together with the SME. Tagging a time to each event allows the tools to create an event timeline once all events have been described.

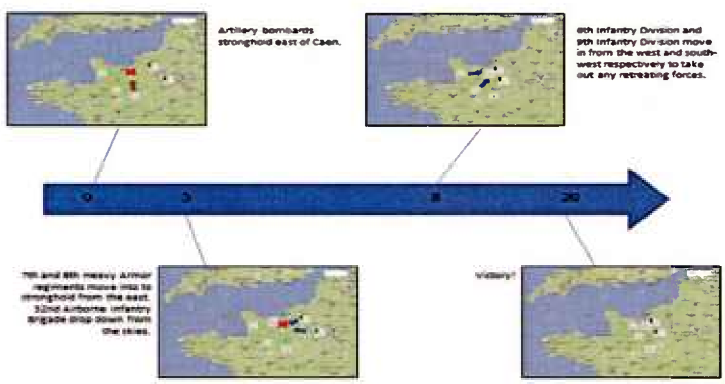


Figure 2. Auto-generated event timeline. The images of the key events will be presented in a timeline together with the notes that were written for each event.

FACILITATION FOR CTA PRACTITIONERS

The prototype tool was used in a series of interviews using CDM in the context of wargaming (Tan, Tee & Soh, 2014). In addition to the prototype tool, CDT was created using a spreadsheet to document the information collected from the interviews. A total of six CTA practitioners were involved in these interviews. At least two CTA practitioners were paired for each interview. The objective of the study was to elicit the warfare knowledge from experienced officers, and translate the knowledge into a building a set of scenario-based training simulator. 40 incidents were captured from the 27 experienced officers.

Accurate representation of events using the scenario drawing and visualization tool

By creating images of the events during the interviews, the CTA practitioners had a better comprehension of the incident described by the SMEs. Misunderstanding by the CTA practitioner was also easily resolved as the SMEs could immediately point out any incorrect representation of the event. As the timeline provided a visual representation of how the events transited, the CTA practitioners could easily identify any gaps between the events that might be critical and probe further.

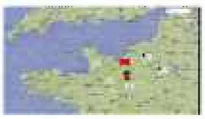
CDT provides affordance on the areas to deepen

Information elicited during the interviews was typed into the respective CDT columns. As decision points were identified, the CTA practitioners sometimes might skip the current set of questions and delve into the identified decision point. The information that hadn't been asked was hence left blank. The empty spaces served as an affordance for the CTA practitioner to identify which area/s required further deepening. Therefore CDT helped by relieving the CTA practitioners from having to remember the questions that he had to ask.

Knowledge easily understood and interpreted

We added an additional column in CDT to include the images of the events created by the scenario drawing tool. During the analysis phase of the interview data, the images certainly helped the CTA practitioners to quickly recall the incidents without having the read through the text. It was also easier for the SMEs to recall what incidents they shared when the CTA practitioners sent the CDT spreadsheet to them for clarification and verification. An example of the CDT is shown in Table 1.

Table 1. Example of Cognitive Demands Table (CDT): The event image created from the scenario drawing tool is added into CDT to better understand the incident. The empty cells in CDT help the CTA practitioners to identify areas that have not been addressed during the interview.

Events	Storyline and Decisions	Strategies	Challenging Cognitive Demands	Expert-Novice Differences	Lessons
	6th and 9th infantry Division move in from the west and south-west respectively to take out any retreating forces.	Adversary starting to retreat eastwards.		Novice might perform X instead because…	(Keywords highlighting the lesson learnt)

Empty cells indicate areas that have not been addressed.

CONCLUSION

The process of conducting CTA interview is laborious and cognitively demanding. With the increasing demands of CTA for expert knowledge elicitation, there is a need to develop tools to facilitate the interviewing process to ensure comprehensiveness and accuracy in the information collected. The scenario drawing and visualization prototype tool aims to facilitate the communication between the CTA practitioner and the SME so that the information described by the SME can be represented quickly in a visual form. Coupled with CDT, the interview process is less cognitively demanding as both tools allow the CTA practitioner to identify topic of interest during the CDM interviews as well as for analysis phase.

REFERENCES

Blessing, S. B., & Anderson, J. R. (1996). How people learn to skip steps.Journal of experimental psychology: learning, memory, and cognition, 22(3), 576.

Chipman, S. F., Schraagen, J. M., & Shalin, V. L. (2000). Introduction to cognitive task analysis. In J. M. Schraagen, S. F. Chipman and V. L. Shalin, eds. *Cognitive task analysis*. Mahwah, NJ : Lawrence Erlbaum Associates, 3-23.

Crandall, B., Klein, G. A., & Hoffman, R. R. (2006). Working minds: A practitioner's guide to cognitive task *analysis*. Cambridge, MA: Mit Press.

Feldon, D. F., Peugh, J., Timmerman, B. E., Maher, M. A., Hurst, M., Strickland, D, Gilmore, J. A., & Stiegelmeyer, C. (2011). Graduate students' teaching experiences improve their methodological research skills. *Science*, *333*(6045), 1037-1039.

Feldon, D. F., Timmerman, B. C., Stowe, K. A., & Showman, R. (2010). Translating expertise into effective instruction: The impacts of cognitive task analysis (CTA) on lab report quality and student retention in the biological sciences. *Journal of research in science teaching*, *47*(10), 1165-1185.

Hoffman, R. R., Coffey, J. W., Ford, K. M., & Carnot, M. J. (2001). Storm-lk: A human-centered knowledge model for weather forecasting. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,45*(8), 752-752.

Klein, G. A., Calderwood, R., & Macgregor, D. (1989). Critical decision method for eliciting knowledge. *Systems, Man and Cybernetics, IEEE Transactions on,19*(3), 462-472.

Militello, L., & Hutton, R. (1998). Applied cognitive task analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics, 41*, 1618-1641.

Radtke, P. H., & Frey, P. R. (1997). Sea Stories: A Collaborative Tool for Articulating Tactical Knowledge. *Presented at Interservice/ Industry Training Simulation, and Education Conference.* Florida, USA.

Tan, L. S. A., Tee, H. G., & Soh, B. K. (2014). A Framework for Knowledge Transfer: Applying the Critical Decision Method for Instructional Design in Wargaming. *Paper presented at the 3rd International Conference of the South East Asian Network of the Ergonomics Society (SEANES)*, Singapore, SG.

Tofel-Grehl, C., & Feldon, D. F. (2013). Cognitive Task Analysis–Based Training A Meta-Analysis of Studies. *Journal of Cognitive Engineering and Decision Making*, *7*(3), 293-304.

The SAnTA Recommender System Aids Naturalistic Decision Making

Frank LINTON[b], Sarah BEEBE[a], Mark BROWN[b], Casey FALK[b], and Mark ZIMMERMANN[b]
[a]*Ascendant Analytics*
[b]*The MITRE Corporation*

ABSTRACT
This paper presents a practice-oriented original contribution that advances the application of NDM. SAnTA is a prototype recommender system that is currently being piloted. SAnTA embodies expert reasoning and supports individual and team performance in a complex real-world setting. SAnTA's users are analysts who create reports for other decision-makers; the reports are based on data and information from multiple sources. Both the data and the decision-makers' needs (the analysts' sources and goals, respectively) may be dynamic, uncertain and continually changing, consequently analysts must also make decisions about their work processes and products under these same circumstances. SAnTA aids these analysts' decisions. Their reports are usually not needed in real-time, but they may be urgent and mistakes can have significant consequences.

KEYWORDS
Decision Making; Education and training; Externalized/Embedded Cognition; recommender system; structured analytic technique; intelligence analysis.

INTRODUCTION

This paper presents a practice-oriented original contribution that advances the application of naturalistic decision making (NDM). The Structured Analytic Technique Advisor (SAnTA) is a prototype recommender system that solicits user input as responses to questions and suggests structured analytic techniques (SATs). SAnTA embodies expert reasoning and supports individual and team performance in a complex real-world setting. SAnTA's users are analysts who create reports for other decision-makers; the reports are based on data and information from multiple sources. Both the data and the decision-makers' needs (the analysts' sources and goals, respectively) may be dynamic, uncertain and continually changing, consequently analysts must also make decisions about their work processes and products under these same circumstances. SAnTA aids these analysts' decisions. Reports and briefings from analysts are needed on time scales ranging from months to minutes and mistakes can have significant negative consequences.

Unlike much NDM research (Klein, 2008), this work does not address decisions made while fighting a fire or a battle, but instead addresses decisions made when creating the analysis that may lead to, or prevent, a battle – or a war. The analyst's life generally is not on the line, but their professional reputation may be. In this environment, the aspects of NDM that apply are making "...tough decisions under difficult conditions such as limited time, uncertainty, high stakes, vague goals, and unstable conditions." (Klein, 2008)

SATs, including qualitative aids aimed at promoting critical thinking and mitigating cognitive mind-sets and biases, are increasingly used by analysts and are mandated by analytic production organizations. SATs are now taught in analytic training programs world-wide. SAT usage is supported by reference materials and by facilitators who guide analysts using a question-based approach to select relevant SATs, and apply them in the resulting analytic process. Yet barriers remain: There are relatively few experienced tradecraft analytic advisors, the number of SATs is growing, and the reference materials offer little guidance on how to select the best SAT for the analytic challenge at hand. Thus, the conditions for improper selection and misapplication of SATs are ripe.

SAnTA is a recommender system intended to prompt critical thought, suggest applicable SATs, and provide further resources and guidance on SATs. SAnTA employs a question-based critical thinking approach to prompt consideration of both process- and problem-related factors that affect relevant SAT selection. The tool provides recommendations on the basis of analysts' characterizations of their progress through the analytic process. The resulting recommendations are ranked and include explanations of the basis for the recommendations and technique descriptions.

SAnTA's end users are essentially characterizing their situation in order to learn what sorts of bias are most likely to be present at this point in the analysis and with which SATs (if any) they can best be overcome. SAnTA's end users are typically less-experienced, and do not have enough expertise to make optimal SAT selections by recognition-primed decisions (RPD) (Klein, 2008).

Recommender systems are frequently found in web-based commerce, where they recommend purchases to consumers. In that domain, much research is focused on algorithms for generating good-enough recommendations quickly. In contrast, very little recommender research has focused on recommendations as decision aids, helping people make the best choices about what to do next, as SAnTA does (Chen, de Gemmis, Felfernig, Lops, Ricci, Semeraro, & Willemsen, 2013; Felfernig, Chen, & Mandl, 2011).

Recommender systems are characterized by how their recommendations are generated (Ricci, Rokach, Shapira, & Kantor, 2011). They are characterized as collaborative filtering (based on user ratings), content-based filtering (based on item attributes), and knowledge-based filtering (explicit knowledge of how certain items meet users' needs). SAnTA is a recommender of this third type, a knowledge-based recommender. In SAnTA, expert knowledge pertaining to the topic of selecting and applying structured analytic techniques (SATs) is acquired, synthesized, and encoded.

The challenge for the SAnTA research then, was to demonstrate a system that a) had the expertise to select SATs appropriate to analysts' situations, and b) to be useful, that is, analysts, trainers, and facilitators would choose to use SAnTA in their work.

METHOD

We assembled a research team with strengths in analytic techniques, knowledge capture, and recommender systems. We began our research process by interviewing representatives from the across the analytic community who were aiming to improve analytic tradecraft quality. We discussed key needs, approaches, opportunities, and constraints. We confirmed the importance of selecting and applying SATs appropriately and our perception that an SAT recommender could help in addressing the problem. While conducting these interviews we also learned that the likely end users would be not only analysts but also facilitators and instructors.

We decided to conduct applied research, and that our success criteria for the design was that it be useable, useful, and used. We prototyped and evaluated various designs for the user interface. We experimented with alternative representations of domain expertise. We enlisted a subject matter expert (SME) who is a former analyst and front-line manager of analysts, an instructor of SAT usage, an author or co-author of several SAT how-to texts, and, it happens, an articulate advocate for the software. For a programmer we hired a summer intern who turned out to be extremely talented.

We created a series of prototypes (using pure JavaScript for ease of transfer to the end users' security-conscious environments) into which our SME input the various factors that influence SAT recommendations, a list of SATs to recommend, and a matrix characterizing how each factor contributed to the strength of the recommendation of each SAT. The SME tested the pre-prototype using historical case studies. We demonstrated this pre-prototype version to the various representatives with whom we had spoken originally, and others, and obtained very positive feedback on our approach. For example:

> *"I have seen many attempts to recommend SATs and they all have failed. This tool works. Let me know how I can help test this in my classes."*

We then partnered with one analytic tradecraft group and supported them as they input their own set of SATs to recommend, their set of factors that influenced their recommendations, and the matrix indicating how each factor contributes to the recommendation of each SAT. The resulting version of SAnTA encapsulates the synthesized expertise of this group in its real-world context. These SMEs are currently checking that SAnTA makes sensible recommendations by giving SAnTA case studies - previously completed projects - and examining SAnTA's output. This version of SAnTA is also being piloted currently with a number of analysts to obtain their feedback. We continue to iterate the software to improve the design of the user interface, the set of factors that influence recommendations, the set of SATs recommended, and the relationships among them. In order to make SAnTA easy to install in the customer's environment initially, we intentionally sacrificed the ability to log user activity - thereby losing the ability to obtain usage analytics as a basis for continuous design improvement. While an instrumented version of SAnTA would be more complex and harder to install and maintain, it would be a significant step forward in the long-term sustainability of the product.

To summarize, the project goal was to develop a prototype SAT recommender system that is usable, useful, and used in real-world analytic environments characterized by uncertainty, limited time, tough decisions, and high stakes. We began by building a prototype and having one SME identify key factors used when selecting SATs. With this as an example, we found a group of SMEs and had them, as a group, validate key factors and, for each factor, the corresponding SATs. We took pains to ensure that our software was modular and that SMEs from different, but related, domains could replace the factors and SATs with their own set. We developed and are distributing an evaluation form to obtain end user feedback. We are asking SMEs to input previous analytic tasks to compare the SATs recommended by SAnTA with the SATs actually used.

RESULTS

For organizations, SAnTA captures corporate knowledge (synthesized domain expertise) in a visible, accessible, updateable, interactive, and user-focused software tool. For decision-making analysts, SAnTA first requires users to articulate their situation in tradecraft terms, that is, to pause and take a meta-level view of their work in responding to SAnTA's questions. Next, it captures their articulated inputs for later introspection and sharing. Third, SAnTA overcomes limitations of recall, working memory, and availability bias, enabling analysts to go beyond the first SAT that comes to mind and select from the most-likely-suitable ones. Choosing which SAT(s) to use from this set of SAT recommendations prompts further reflection and discovery.

SAnTA inverts the dynamics of selecting a structured analytic technique. In the past, analysts first attended a course or read some material about SATs in a book or online. Then, when the need for an SAT arose, the analyst would recall the SATs mentioned or review a list of SATs, and try to select one that seemed suitable. If all else failed, they could attempt to obtain help from a facilitator. In contrast, with SAnTA, analysts simply describe their situation to SAnTA; characterizing where they are in the analytic process, the resources and constraints in their environment, and so on, and SAnTA provides the user with a small set of recommended SATs, together with a rationale for their selection, the strength of its recommendation, a visualization of where in the analytic process each one applies, and links to further tools or information.

To date SAnTA research has accomplished several things: first, it documents corporate knowledge and best practices (by requiring their articulation and synthesis), and it does so in a software system that makes this knowledge easily accessible by analysts, trainers, and facilitators. Next, SAnTA contains, and can recommend when appropriate, more SATs than any single facilitator. Third, even if it should turn out to be the case that users need not apply an SAT, in the process of determining this, they will have reviewed a number of questions that every analyst should consider when creating a product. Finally, SAnTA has been developed in a manner that makes it easy to edit or change out the knowledge base it uses.

To date we have learned that first, when users input data from their past cases, the recommendations are reasonable. Next, ten to twenty questions appear to be sufficient to capture the information needed to make good recommendations (we expect this number could be reduced with further analysis). Third, the current software is easily transferred to and is working well in the analysts' environment. Finally, while the software is useful as-is, numerous improvements and refinements could be made.

At this point we have buy-in from representatives of the user community. That buy-in has enabled us to obtain funding to transfer the technology to the analytic community and funding to continue the research in a different domain where the approach is also likely to prove fruitful. We have paid attention from the start to creating a product that will be useable, useful, and used, and the feedback we have obtained so far indicates an increasing likelihood of achieving that goal. We have created a representation of domain expertise - for domains with certain characteristics - that makes it relatively easy to capture, synthesize and make that expertise readily accessible to a broad audience. We have a system that is a shell capable of doing the same for any similarly structured expertise. We have created a system that is expected to help analysts think more thoroughly, by making SAT options visible at the time of their application.

DISCUSSION

The challenge for the SAnTA research was to demonstrate a system that a) had the expertise to select SATs appropriate to analysts' situations, and b) to be useful, that is, analysts, trainers, and facilitators would choose to use it. Obtaining initial input from the community of users across multiple organizations helped ensure the very first design would be useful, and input to the pre-prototype from two SMEs illustrated the nature of the desired expertise. At this point SAnTA's expertise and utility have been demonstrated to various members of the analytic community and their anecdotal response strongly favors SAnTA's design. When demonstrated to a group of trainers and facilitators SAnTA was seen as useful. That group of domain experts was easily able to express its synthesized expertise in SAnTA's representation. That representation is currently being checked by them using previously completed analytic products.

One of SAnTA's significant strengths is its user orientation. Instead of thumbing through pages of SATs looking for one whose 'when to use' attributes match their situation, analysts describe their situation to SAnTA and

SAnTA generates a set of recommended SATs, together with supporting information that makes it easy for analysts to match the most appropriate technique to their analytic challenge.

Limitations of this report

1. As of this writing, January 2015, actual use by analysts is in progress.
2. As of this writing, January 2015, use by more than one group in more than one organization is in progress.

Further research on SAnTA

1. Validate the content (factors, weights, and recommendations) using examples from the past.
2. Develop a more-formal means of validating content, especially recommendations.
3. Determine the minimum factors to consider for generating valid recommendations.
4. Find optimal visualizations or ways to present SAnTA's recommendations.
5. Instrument SAnTA and gather usage data to improve the design.
6. Study the usage of SAnTA and the factors influencing the use or non-use of SAnTA.
7. Make it as easy as possible to edit, and to swap in/out, sets of domain knowledge.

Implications

The use of recommender systems for learning (versus consumption) is a neglected area of research. Recommender systems that support informal workplace learning appear to be a potentially valuable tool for the collaborative bootstrapping of expertise. SAnTA demonstrates that recommender systems can support less-experienced naturalistic decision-makers, such as analysts, in specific circumstances.

The conditions of NDM that apply when doing analysis, i.e., making "...tough decisions under difficult conditions such as limited time, uncertainty, high stakes, vague goals, and unstable conditions" (Klein, 2008) make it difficult to pause and think at a reflective or metacognitive level. The presence of an SAT recommender may take pressure off the analyst in that respect, providing quick solutions to the task of selecting suitable, defensible, SATs for use. In these circumstances, giving analysts control, i.e., the ability to change their responses and observe the resulting change in recommendations, together with seeing reasons for the change, enable them to evaluate the specificity, generality, and robustness of the recommendations, and to understand the applicability conditions of various SATs, and to begin to internalize this knowledge for later use. The ability of an analyst to save a SAnTA session, including their inputs and the resulting recommendations, enables both users and those reviewing their work later to reflect on these perceptions and decisions, and may also lead the analyst toward 'just knowing' which SAT to select.

To summarize, SAnTA is a knowledge-based recommender system that solicits user input as responses to questions and suggests most-likely-relevant SATs. The user must make the decision of which SAT(s) to use, if any. SAnTA (Figure 2) provides brief descriptions of each SAT, explanations of how the user's responses influenced the SAT recommendations, and links to further information on each SAT. SAnTA is written in JavaScript, a design decision that makes it easy to port the software to many environments. SAnTA does have hooks for instrumenting the application and logging how it is used; a future capability that will require SAnTA be linked to a server, which we are currently avoiding for ease of portability. The recommendation algorithm is a simple one. To start with, SMEs rate the applicability of each SAT to each response creating a matrix. As users enter their responses, SAnTA computes the ranking of each SAT based on the current response set. As more responses are entered, SAnTA becomes more certain of its recommendations. This design means SAnTA is easily customized. Sets of SATs and sets of questions and responses are easily exchanged.

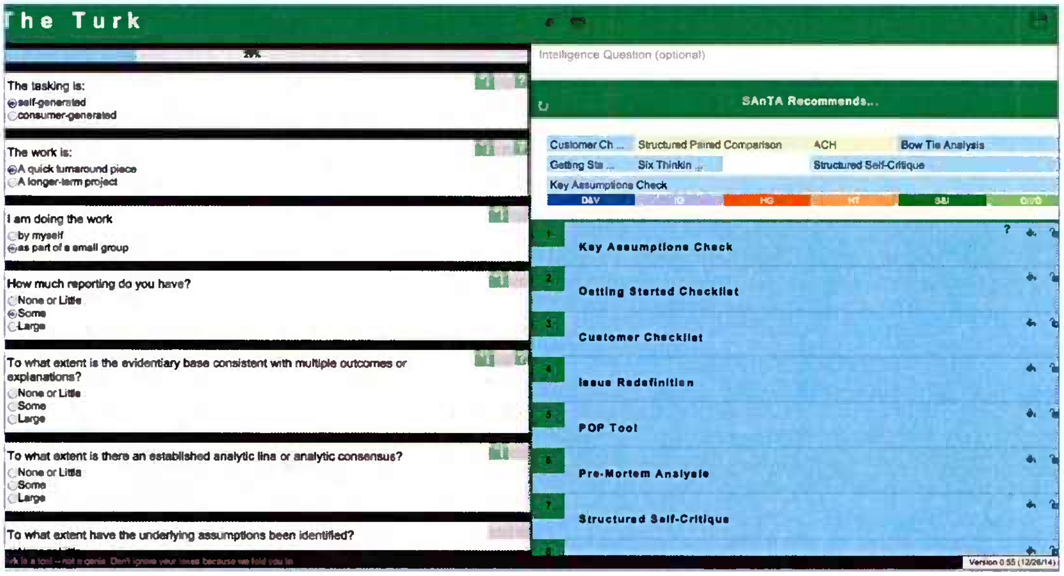


Figure 2. The SAnTA user interface (here: 'The TURK'). In this screenshot, the analyst has begun to describe her analytic situation to SAnTA by responding to the questions on the left side of the screen. On the right side of the screen SAnTA has begun to respond by returning a ranked list of SATs and a graphic depicting where in the analytic process the SATs apply. Not shown are popups describing each SAT, SAnTA's confidence in its recommendations, and popups explaining icons, abbreviated text, etc.

REFERENCES

Chen, L., de Gemmis, M., Felfernig, A., Lops, P., Ricci, F., Semeraro, G., & Willemsen, M. 2013. Workshop on human decision making and recommender systems: Decisions@RecSys'13. RecSys'13, October 12–16, 2013, Hong Kong, China. ACM 978-1-4503-2409-0/13/10. http://dx.doi.org/10.1145/2507157.2508002.

Felfernig, A., Chen, L., & Mandl, M. 2011. RecSys'11 Workshop on Human Decision Making in Recommender Systems. RecSys'11, October 23–27, 2011, Chicago, Illinois, USA. ACM 978-1-4503-0683-6/11/10.

Klein, G. Naturalistic Decision Making. Human Factors, Vol. 50, No. 3, June 2008, pp. 456–460. DOI 10.1518/001872008X288385.

Ricci, F., Rokach, L., Shapira, B., & Kantor, P. (Eds.) (2011). Recommender systems handbook. New York: Springer.

What is the Story Behind the Story?

Two Case Studies of Decision-making under Stress

Anne-Claire MACQUET[a], Héloïse LACOUCHIE[ab],
[a]National Institute of Sports, Expertise and Performance (INSEP), Paris, France
[b]René Descartes University, Paris, France

ABSTRACT

This study aimed to explain how judokas made their decisions under stress. Two female elite judokas participated in the study. An interview was conducted with each athlete separately. Athletes were asked to describe and comment on the decisions made under stress during a recent important match. In reference to the Critical Decision-Making method, interviews were used to enable judokas to describe the decisions made during the match, the processes used to make these decisions and the effect of these decisions on the match. Results showed that judokas used a situation recognition process to make decisions. Results also showed that decisions prioritised the use of judokas' favourite techniques. Situations in which favourite techniques could be used arose in one of two ways: either the situation arose naturally, or judokas decided to manipulate the situation to create the conditions required to implement their favourite technique; then they carried it out.

KEYWORDS

Decision-making; Expertise; Elite sports; Uncertainty management.

INTRODUCTION

In high-level sport, the ability to make efficient decisions under stress is a key factor in performance. Salas, Driskell and Hughes (1996) defined stress as "a process by which certain environmental demands... evoke an appraisal process in which perceived demands exceeds resources and results in undesirable physiological, psychological, behavioral or social outcomes" (p. 6). In dynamic and complex situations such as in sports (e.g., Macquet, 2009) and the military (e.g. Cannon-Bowers & Salas, 1998), characteristics which can be considered as stressors appear in an environment demonstrating: (a) numerous information sources; (b) incomplete and conflicting information; (c) rapid change in the situation; (d) adverse physical conditions; (e) high stakes and performance pressure; (f) intense time pressure; (g) high information load; (h) interference; and (i) threat. In sports, Woodman and Hardy (2001) differentiated between stressors coming from outside the sport organization (e.g., disturbances from friends) and those that are the consequences of the organizational climate (e.g., sport politics). They defined organizational stress as the "stress which is associated primarily and directly with an individual's appraisal of the structure and functioning of the organization within which he/she is operating" (p. 208). Fletcher and Hanton (2003) studied the sources of organizational stress; they identified four categories of stressors: (a) environmental issues; (b) personal issues; (c) leadership issues; and (d) team issues. These sport studies focussed on stressors related to the competitive environment whereas studies about ergonomics and sport cognitive ergonomics focussed on stressors related to the situation. In order to study decision-making under stress, it is important to focus on the stressors related to the situation, such as those presented by Cannon-Bowers and Salas (1998).

Individuals vary in their sensitivity to stressors (e.g., Gaillard, 2008). The ability to remain concentrated under time pressure, fatigue and threat contributes to the determination of stress tolerance. One of the aims of sport training consists of training under such conditions to improve the ability of athletes to make and implement efficient decisions in uncertain environments. From a scientific and practical perspective, one of the challenges is to understand how elite athletes make decisions in these situations of stress and high stakes.

For many years, a field of research has stressed the need to study decision-making in naturalistic environments in order to conduct meaningful research (e.g., Zsambock & Klein, 1997). One of the major theories in Naturalistic-Decision Making (NDM) is the recognition-primed decision (RPD; Klein, 1997; Klein, Calderwood, & Clinton-Cirocco, 1986). According to this theory, experts use their experience to make good decisions without comparing the multiple alternative courses of action. They assess a situation by comparing it with similar previously experienced situations associated with a typical action and stored in the memory. If the current situation is analogous to the typical situation in the memory, then experts implement the corresponding action and adapt it to the current situation. To assess its potential effectiveness, they mentally simulate its. According to the RPD model, individuals recognize its typicality through salient features that experience has shown to be useful. Recognition has four by-products: (a) expectancies; (b) salient features; (c) plausible goals; and (d) typical action. Results showed that experts determined only a small number of options (Johnson & Raab, 2003; Klein, Wolf, Militello, & Zsambok, 1995: Macquet, 2009), and most often, they determined only one (Yates, 2001).

There are three levels involved in evaluating the situation in a changing context: simple match, diagnose the situation, and evaluate a course of action. In level 1, the situation is rapidly perceived as typical, so the expert can quickly implement a course of action that corresponds to this typical situation. In level 2, the situation is not initially perceived as typical so the expert must clarify it by assessing it to determine its typicality. Then, he/she can carry out a course of action based on the appropriate typical action. In level 3, the expert perceives the situation as typical. Before undertaking a typical action, he/she considers several possible actions and assesses them using mental simulation to determine whether they are likely to work. When he/she considers that one will work, he/she carries it out. However, experts devote more time and energy to assessing what is happening rather than comparing several courses of action.

Previous studies showed the role of practice in decision-making in simulated or standardized situations (e.g., Raab, 2002), and the use of procedural rules to make decisions in competitive situations (e.g., McPherson & Kernodle, 2003). Other studies showed that in a competitive environment athletes reproduced actions that had been practised extensively during training sessions (e.g., Macquet, 2009; Macquet, Eccles, & Barraux, 2012; Macquet & Fleurance, 2007). The NDM and RPD models could provide a theoretical and practical tool for understanding decision-making in competitive situations for several reasons. NDM focuses on real settings involving complex problems, intense time pressure and high stakes, such as sport competitions. The RPD model could provide theoretical and practical insights into athletes' use of their experiences during high-stake competitive situations. It could also provide insights into the number of decisions made by athletes in competitive situations. Furthermore, it could extend knowledge about information perceived during the course of action and reported by athletes.

The present study aimed to explain how judokas made decisions under stress during important competitions. More specifically, it aimed to explain the cognitive experience, knowledge and expertise of judokas from a decision-making perspective.

METHOD

Participants

Two female judokas, aged 20 and 32 years, volunteered to participate in the study. They had been practising judo for seven and 19 years respectively. They were ranked within the top 50 judokas worldwide at the time of the study. The first athlete had been competing at the top level for five years and the second athlete for 14 years. They had won medals at the European and World Championships. The second athlete had also won a World Championship. The athletes were informed of the purpose of the study and assured of anonymity. The study was approved by a local ethics committee. Athletes were given the pseudonyms A1 and A2 to provide some degree of confidentiality.

Data Collection

Two interviews were conducted with the athletes separately, in reference to the critical decision-making method (Crandall, Klein, & Hoffman, 2006). The Critical Decision-Making (CDM) method involves an intensive interview. It consists of asking an individual to comment on an incident he/she experienced and the difficult decisions he/she made. The interviewers invite the interviewee to give information about decision-making and sense-making while recalling a specific incident. Interviewers try to probe progressively deeper on cognitive issues by asking the interviewee to comment on details, background influences and tactics in relation to the decision made. The research team has to understand the story of the specific event and the cognitive demands of the task and setting. Two interviewers are required. The first one conducts the interview, acts as a facilitator and takes notes. The second one takes notes and is responsible for keeping track of the overall interview progression.

The interview comprises four steps: (a) incident identification; (b) time-line verification; (c) deepening, and (d) "What If" queries. The first step consists of selecting an appropriate incident for in-depth examination. This incident needs to contain non-routine and challenging events. It also needs to have happened during the month preceding the interview, in order to enable the participant to remember the difficult situations and decisions easily. Such an incident enables the researcher to cover elements of expertise and related cognitive phenomena. It also enables to learn about the components that characterize skilled performance and expertise. Once the incident has been selected by the interviewee, the interviewer asks the interviewee to give a brief account of the match. This account provides the foundation for the remainder of the interview.

The second step consists of having a clear, defined and verified overview of the incident, by eliciting key events and segments. The interviewee comments on events in greater detail when he/she relives the event in his/her mind. This step is key. Recalled information enables the interviewer to structure the interview and provides information from which to construct a timeline. The timelines of the interviews were used to better understand the decisions made and their effects on the outcome of the contest. Due to text length constraints, they are not presented within the results of the study.

The third step consists of probing. This enables researchers to see "inside the expert's head and look at the world through his or her eyes" (Crandall et al., 2006; p. 77). This step consists of going beyond the facts of the incident to elicit the participant's perceptions, expectations, goals, tactics and the consequences of decisions made. It allows researchers to construct a detailed account of the incident from the interviewee's point of view and to provide the story behind the story, namely the cognitive experience, knowledge and expertise.

The fourth step relates to "What If" queries. "What if" questions allow researchers to see differences between experts and novices and possible vulnerabilities. The interviewee is invited to consider how his/her decisions might have been different if he/she were a novice.

Judokas described a match that lasted 5 min 15 seconds for A1 and 1 min 30 seconds for A2. The interviews lasted 35 and 45 min respectively. Interviews were recorded and transcribed in full.

Data Processing

Data processing consisted of explaining, "What is the story behind the story?" What are the data saying that I do not yet know? Data processing was done using the constant comparative method (Corbin & Strauss, 1990). This method consists of identifying a phenomenon of interest, and a number of local principles or process features of the phenomenon of interest, and categorizing the data based on the initial understanding of the phenomenon. Two researchers analysed the verbal reports separately.

Data processing involved three steps: (a) identification of meaningful units in relation to tough decisions and their effects on the outcome of the contest; (b) construction of the timeline of decisions taken and their effects on the match; and (c) identification of the elements used to make decisions in relation to the RPD model. In the first step, researchers identified the meaningful units in relation to decisions made, and perceptual and cognitive components of decisions in relation to the RPD model. The second step consisted of the construction of the timeline of the match, namely the decisions made and their consequences on the way the situation evolved and the score. It also enabled researchers to differentiate between the tough decisions during the match and their effects on the score, and the decisive decisions that caused the match to end. Finally, it allowed researchers to build up the story of the match. The third step enabled researchers to identify the cognitive components of the decisions made in relation to the by-products of the RPD model (Klein et al., 1986). It also allowed the level of the RPD model to be identified: (a) simple match; (b) diagnosis; or (c) mental simulation.

After each data processing step, data were constantly compared until saturation was reached, which occurred when no further meaningful units and categories were identified from the data. The researchers compared their results and discussed any initial disagreement until consensus was reached. Interview transcripts were divided into 330 meaningful units.

RESULTS

The matches reported by the contestants presented high stakes: A1 was aiming to win the French Championship and A2, third place in the World Championship. A1 had lost against her current opponent in a previous competition. Story building and timelines of the matches allowed the results to be presented in two parts: (a) a succession of ineffective decisions that led to a tight score; and (b) decisive and effective decisions that caused the match to end. In each part of results, decisions were compared with the RPD model.

Succession of Ineffective Decisions that Led to a Tight Score

During elite judo training, judoka learn many judo skills and train to improve these so that they become routine skills that judokas can implement quickly and efficiently. A judoka and his/her coach choose one of these skills, depending on the athlete's physical characteristics and preferences. Judokas practice this one more extensively than the others. This specific skill is called the favourite technique. During a contest, each judoka tries to use his/her favourite technique in order to impose his/her way of competing on the opponent and win the match. Conversely, the opponent tries to prevent the judoka from implementing his/her favourite technique and to use his/her own favourite technique. In judo, the way athletes stand in front of an opponent and grip their opponent's kimono (i.e., kumikata) largely determines the technique the athletes are going to implement and the probability of achieving a positive outcome. Each judoka has his/her own kumikata depending on whether the judoka is right or left-handed, tall or small and so on.

Results showed that participants and their opponents both tried to implement their favourite technique on occasions without success. For example, A1 said: "During the first part of the contest, my second seois [participant's favourite technique] didn't work, the referee judged it to be a false attack and sanctioned me twice. Seoi is my favourite technique." Results also showed that for both judokas some decisions made were ineffective at winning the match, or led to penalties. This succession of ineffective decisions lasted 5 min (normal time of the contest period) for A1 and 1 min 30 seconds for A2. At the end of these periods, participants reported difficulties in selecting effective decisions. For example, A1 said: "The match was very tight. We both had penalties and the score was equal at the end of regular time."

Categorization of the elements used by the participants to make decisions during the first period of the match was made according to the four by-products of the RPD model: (a) expectancies, (b) relevant cues, (c) plausible goals, and (d) typical action. Results showed that expectancies related to what they thought the opponent would do. They involved anticipation of a specific action that the opponent might carry out in view of an opponent's or participant's abilities and tendencies, level of expertise in a specific situation and experience with this opponent in current or previous contests. For example, A2 said: "She's got a high grip [she catches the kimono by its upper part] and she's aggressive while on the attack. Results indicated that relevant cues related to: the opponent's freshness and involvement, the participant's freshness and involvement, and the score. For example, A1 said: "After three minutes, I noticed that she was tired, she had less strength, her reactions were slower. I was tired too but not as tired as she was." Plausible goals were seen to consist of the number of goals and decisions the participant considered she could implement in the course of the action. For example, A2 said: "I was focussing on this sleeve I had to catch". Participants reported only one goal at a time. Results demonstrated that typical actions referred to the actions that were often undertaken by the opponent or participant in a typical situation, and more specifically, the favourite technique. They referred to an association between a condition and an action to be carried out. The players compared the current situation and event to prior ones. This comparison led them to recognise the situation as typical. They then implemented their favourite technique directly. For example, A2 said: "As she isn't a real right-hander, I prevent her from moving her hand up while gripping my kimono and hold her kimono at the top."

Results also indicated that participants' decision-making in this part of the contest related to the first level of the RPD model: participants recognized the situation rapidly and implemented a course of action.

At the end of this part of the match, athletes commented on the ineffectiveness of the decisions they made and the effectiveness of the opponent's decisions. For example, A1 said: "The match was very tight, no one had an advantage. There was no fall, no real attack." In another example, A2 said: "At this moment of the match, she gripped my kimono first. I was in danger, not enough to be attacked, but I was in danger at the grip level."

Decisive and Effective Decisions that Caused the Match to End

The second part of participants' matches was preceded by a pause to allow the athletes to return to the centre of the contest mat and adjust their kimonos. This pause also allowed them to adapt their initial technique to the unfolding events. For example, A1 said: "She was taller than me and right-handed. I had to control her arm and move her. I had to let her start the action. I told myself to put more pressure on her, she's more tired than I am. I must remain alert, too." In another example, A2 said: "I must move my hand up to dominate more and move her forward."

Participants started the second part of the match and rapidly made decisions that led to them winning the match. These decisions were decisive. Results showed that their decisions related to the participants' favourite techniques. However, they were different to participants' earlier decisions involving favourite techniques. This time, participants' decisive decisions involved action-reaction. Participants made a first decision in order to create appropriate conditions for the effective implementation of their favourite techniques (i.e., second decision). In other words, participants did not change their decision to implement their favourite technique. Rather, they changed the situation to make it possible to implement their favourite technique. For example, A1 pretended to implement a routine forwards but instead, implemented this routine backwards. A2 made her opponent free her hold on the kimono to force her to react. Results showed that in order to enable them to undertake their favourite routine, participants made a decision aimed at manipulating the situation to suit the technique they wanted to implement. In other words, instead of changing their decision totally, they decided to change the situation so that they could implement their favourite technique effectively.

Results showed that the decisive decisions that led to the matches ending concerned the four by-products of the RPD model. Participants reported on expectancies while commenting on both opponents' abilities and tendencies and their own abilities and tendencies. For example, A1 said: "She is a very explosive girl during the three first minutes. Then she slows down and is less explosive while attacking". Participants also reported on relevant cues related to opponents' freshness and involvement. For example, A1 said: " She breathed more often, she had difficulty breathing." Results indicated that participants had only one plausible goal at a time and also that participants' typical actions related to their favourite techniques and preceding events. For example, A2 said: "I pulled her forwards as though I was going to implement my favourite routine forwards but I pushed her backwards. I tried to surprise her to make her react backwards, to make her put her weight backwards. I then changed axis and implemented Ippon-Ko" [using a small leg hook and controlling the shoulder]. This decision made me win" (win by ippon).

Finally, judokas reported differences between themselves (i.e., experts) and novices in such situations. Participants commented on their ability to maintain intense concentration and rigor despite fatigue and a tight score for A1, and negative emotions in relation to her failure in the semi-final preceeding her current match for A2 and specific style of A2's opponent. For example, A2 said:

"She has a very specific style. She moves forward all the time, catches the kimono, releases it, catches it, then releases it and so on. A novice is in danger of losing concentration. Instead of concentrating on himself/herself, he/she will be tempted to focus on the opponent. And that's how you lose".
Then A2 said: "I remained concentrated on me on what I had to do to win".

DISCUSSION

The objective of this study was to gain an understanding of the process of decision-making under stress. In order to meet this objective, the researchers used inductive and deductive analysis to understand in detail the difficult decisions made by expert judokas. Results are discussed in two parts: (a) the consistency of the results with the RPD model; and (b) the story behind the story.

Consistency of the Results with the RPD Model

As the RPD model predicts, the results of this study showed that judokas' decision-making was based both on a process of recognition of a typical situation and the use of associations between a typical situation and a typical action. The unfolding situation was compared to a typical situation in the memory. The process of recognition of the situation enables the individual to assess the situation. This process was based on four by-products: (a) relevant cues, (b) expectancies, (c) plausible goals, and (d) typical actions. The judokas perceived relevant cues from visual perception (e.g., focusing on the opponent's sleeve), kinesthesic perception (e.g., feeling a reduction in the strength of her opponent while being gripped by the kimono), and auditory perception (e.g., hearing her opponent's breathing quicken). To our knowledge, previous studies on decision-making in sports (e.g., Macquet, 2009) and other contexts (e.g., Klein et al., 1986) have not reported on relevant cues concerned with the senses. Decision-makers have frequently reported solely on visual data. In some situations, such as in judo, other senses provide important information to assist decision-making. As Macquet et al. (2012) suggested, exploring the role of the different kinds of perception on decision-making would be a worthwile avenue for future research in sports and other contexts.

It can be seen from the results that judokas reported on expectancies based on previous experience with the opponent and knowledge about the opponent's tendencies. These results are consistent with previous research on decision-making in sports based on the RPD model (e.g., Cardin, Bossard, Buche, & Kermarrec, 2013; Macquet 2009).

Results also showed that judokas reported only one goal, meaning that their decision referred to level 1 of the RPD model. This recognition process seemed to depend on their experience of judo situations. Results also suggest that judokas did not have enough time to diagnoze the situation and assess the effectiveness of a possible course of action before implementing or changing it. In judo, because contestants are very close to each other and actions are very fast, contestants must assess the situation rapidly in order to implement an effective decision. Exploring the distance between opponents and time available to act would be a worthwile avenue for future research on expert decision-making in sports. From a practical perspective, assessing a situation rapidly and making a timely decision is a determining performance factor in which coaches must train athletes.

Results indicated that typical actions reported by the judokas mainly concerned judokas' favourite techniques This suggests that judokas compared the unfolding situation to typical situations contained in their memories. When the unfolding situation and typical situation were similar, judokas implemented the typical action appropriate to the favourite technique, and adapted it to the current situation. If not, they waited for the situation to occur.

The Story Behind the Story

In the first part of their contest, judokas implemented their favourite techniques,with negative outcomes: penalties for A1 and turnover for A2. In the second part, judokas carried out two successive decisions ruled together. The first one aimed to change the unfolding situation so that it ressembled the situation in the memory associated with the typical action related to their favourite technique. Once the second unfolding situation matched the one they were hoping to achieve, judokas undertook their favourite technique. In other words, when judokas could not undertake their favourite technique directly because the current situation did not allow it, they manipulated the situation until it matched the typical situation associated with favourite technique. Decision-making stopped when the situation matched the typical situation contained in the memory and associated with the favourite technique, and the typical action fitted the unfolding situation.

These results suggest that decisions could be driven by by two temporalities: immediate and anticipated. In the first part of the match, judokas expected that their decisions would enable them to score points. In the second part, they expected their inital decision to enable specific changes in the situation, which would provide the required conditions to implement their favourite technique and score points. In the first case, the decision was a one-off decision; in the second case, it was a two-stage decision. It can be expected that such decisions are made under stress solely by expert judokas. Exploring differences in decision temporalities could be a worthwhile avenue for future research.

These results also suggest the high level of expertise of the judokas. Under intense time pressure, stress and fatigue, judokas were sufficiently alert to assess the situation and make a two-stage decision.

Results showed that the motivation of A1 changed at the end of the first part of the contest. The perception that her opponent's fatigue had increased drove A1 to pursue her strategy. Although the score was tight, A1 was not as tired as her opponent. This difference in perceived fatigue enabled A1 to feel capable of winning the match. This result suggests that A1 had psychological momentum. Psychological momentum means a power that increases in relation to the decrease in the opponent's involvement (e.g., Gernigon, Briki, & Eykens, 2010). This power changes the perception of oneself and the opponent, influences the belief that the result will be success, and improves or maintains involvement in the match. In this study, results indicated that this psychological momentum was positive, meaning it enabled A1 to maintain and improve her performance. It was related to the perception of the opponent's strength when the opponent tried to grip A1's kimono. This momentum enabled A1 to feel that it was possible to win the match, even though the score was tight. It brought A1 closer to victory and made her more involved.

Finally, results showed that differences between experts and novices in such situations related to the ability to maintain concentration and rigor thoroughout the matches, despite fatigue and negative emotions. This suggests a perspective for training: coaches might train athletes under conditions of fatigue and negative emotions to prepare them for stress and fatigue in competitions.

This study presents some limitations. Firstly, it did not feature other contestants and contests for comparison. The extent to which decision-making under stress in contest sports relates to a recognition process and an action-reaction plan is therefore unknown. There are very few studies on elite athletes, and sport psychology research often involves few participants because only a small number of athletes reach elite level (e.g., Macquet & Kragba, in press).

CONCLUSION

The data tend to support the view that the decisions of expert judokas under stress were based on recognition and experience. Judokas used their favourite technique first. As their decisions proved ineffective, they manipulated situations to make them resemble the typical situations stored in their memories and then implemented the corresponding typical actions, namely their favourite techniques. The continued study of athletes' decision-making under stress will improve our understanding of cognitive processes and performance.

ACKNOWLEDGMENTS

The authors would like to thank the athletes for their participation in this study.

REFERENCES

Cannon-Bowers, J. A., & Salas, E. (1998). Individual and team decision-making under stress: Theoretical underpinnings. In J. A. Cannon-Bowers & E. Salas (Eds.) *Making decisions under stress* (pp. 17-38). Washington, DC: American Psychological Association.

Cardin, Y., Bossard,C., Buche, C., & Kermarrec, G. (2013). Investigate naturalistic decision-making of football players in virtual reality environment: Influence of viewpoints in recognition process. Proceeding of the Eleventh International Naturalistic Decision-Making Conference, Marseille, France

Corbin, J., & Strauss, A. (1990). *Basics of qualitative research: Grounded theory. Procedures and techniques.* Newbury Park, CA: Sage.

Crandall, B., Klein, G., Hoffman, R. R. (2006). Working minds. Cambridge, Ma: MIT Press.

Fletcher, D., & Hanton, S. (2003). Sources of organizational stress in elite sports performers. *The Sport Psychologist, 17*, 175-195.

Gaillard, A. W. K. (2008). Concentration, stress and performance. In P. A. Hancock & J. L. Szalma (Eds.), *Performance under stress*, (pp. 59-75). Adlershot, UK: Ashgate.

Gernigon, C., Briki, W., & Eykens, K. (2010). The dynamics of psychological momentum in sport: The role of ongoing history of performance patterns. *Journal of Sport and Exercise Psychology, 32*(3), 377-400.

Johnson, J. G, & Raab, M. (2003). Take the first: Option-generation and resulting choices. *Organizational Behavior and Human Decision Processes, 91*, 215-229.

Klein, G. (1997). The Recognition-Primed Decision model: Looking back, looking forward. In C. E. Zsambok, & G. Klein (Eds.), *Naturalistic Decision-Making* (pp. 285-292). Mahwah: Erlbaum.

Klein G. A. , Calderwood, R., Clinton-Cirocco, A. (1986). Rapid decision-making on the fireground. Proceedings of the Human Factors and Ergonomic Society 30th Annual Meeting, 1, 576-580.

Klein, G., Wolf, S., Militello, L., & Zsambok, C. (1995). Characteristics of skilled option generation in chess. *Organizational Behavior and Human Decision Processes, 6,* 63-69.

Macquet, A. -C., (2009). Recognition within the decision-making process: A case study of expert volleyball players. *Journal of Applied Sport Psychology, 21*,64-79. **doi:** 10.1080/10413200802575759.

Macquet, A. –C., Eccles, D. W., & Barraux, E. (2012) .What makes an orienteer an expert? A case study of a highly elite orienteer's concerns in the course of competition. *Journal of Sport Sciences, 30*, 91-99. doi:10.1080/02640414.2011.617774

Macquet, A. -C., Fleurance, P. (2007). Naturalistic decision-making in expert badminton players. *Ergonomics, 50*(9), 1433-1450. **doi**: 10.1080./00140130701393452

Macquet, A. -C., & Kragba, K. (in press).What makes basketball players continue with the planned play or change it? A case study of the relationships between sense-making and decision-making. *Cognition, Technology and Work.*

McPherson, S. L., & Kernodle, M. W. (2003). Tactics, the neglected attribute of expertise. In J. L. Starkes, & K. A. Ericsson (Eds.) *Expert Performance in Sports* (pp. 137-168). Champaign, IL: Human Kinetics.

Raab, M. (2002). T-ECHO: Model of decision-making to explain behavior in experiments and simulations under time pressure. *Psychology of Sport and Exercise, 3*, 151-171.

Salas, E., Driskell, J. E., & Hughes, S. (1996). Introduction: The study of stress and human performance. In J. E. Driskell, & E. Salas (Eds.) *Stress and human performance* (pp. 1-45). Mahwah NJ: Erlbaum.

Von Someren, M. W., Barnard, Y. F., Sandleberg, J. A. C. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. London: Academic Press.

Woodman, T., & Hardy, L. (2001). A case study of organizational stress in elite sports. *Journal of Applied Sport Psychology, 13* 207-238.

Yates, J. F. (2001). "Outsider: Impressions of naturalistic decision-making. In E. Salas, & G. Klein (Eds.), *Linking expertise and Naturalistic Decision-Making* (pp. 9-33). Mahwah, NJ: Erlbaum.

Zsambok, C. E., & Klein, G. (Eds.) (1997). *Naturalistic Decision-Making.* Mahwah, NJ: Erlbaum.

Lifeguard decisions: Naturalistic decision making meets multi-objective optimization on the beach

Joachim MEYER[a] and Daniel HARTMANN[b]
[a]*Department of Industrial Engineering, Tel Aviv University, Tel Aviv 69978, Israel (email: jmeyer@tau.ac.il)*
[b]*Safety Engineering and Management, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel*

ABSTRACT
Lifeguards at beaches continuously solve dynamic decision problems when allowing people to enjoy the water while minimizing the probability of emergency events. We show that this decision meets the criteria for a naturalistic decision making decision problem, and we analyse the decision as a dynamic, real time decision in which factors, such as the number and characteristics of the beach users, the existing physical conditions and the number of lifeguards at a particular station determine the lifeguards' decisions when and how to interfere. In this particular case it is actually not too difficult to define a normative solution for the decision problem. We suggest that normative, quantitative models of NDM problems are possible and can have definite value.

KEYWORDS
Decision making; multi-objective; water safety; beach safety; drowning prevention; life guards.

INTRODUCTION

Naturalistic decision making (NDM) has developed as one of the major theoretical approaches in decision making research (Klein, 2008). It originated from the discontent with much existing work on decision making that failed to address the intricacies of decisions in dynamic real life situations. It posits that it is unrealistic to assume that decision makers intuitively solve complex sets of equation to decide which of a number of alternatives to choose.

However, under certain conditions it is possible to specify a formal model of a decision problem that would be considered an NDM problem. In this paper we will describe such a problem – the decision making of life guards on a Mediterranean public beach, and we present the outline of a possible formal approach to address this decision problem. We obviously do not claim that lifeguards actually solve the equations that need to be used to solve the problem. The optimal solution of such problems is inherently complex and requires long and lengthy computations. Rather, the intuitive solution the lifeguard uses should ideally approach the optimal computed solution, and it may be possible to compare predictions from the optimal solution to the lifeguards' decisions.

Lifeguard decision making

According to the Center of Disease Control's Web-based Injury Statistics Query and Reporting System (http://www.cdc.gov/injury/wisqars), in the U.S. drowning is the most common cause for unintentional injury deaths for ages 1 to 4, is ranked second for ages 5 to 14, is ranked third for ages 15 to 34 and is ranked fourth for ages 35 to 54. Drowning can occur at practically any body of water, but drowning events during the summer months are particularly likely at seaside or lake beaches, especially in states and countries with temperate climates and a culture of beach swimming. Particularly risk factors are male sex, age less than 14, alcohol use, low income, risky behaviour, and lack of supervision (Szpilman et al., 2012).

The major precaution for preventing accidental drowning is to swim at a beach with lifeguard supervision. The increased "water safety" of these beaches is not due to the lifeguards' ability to perform dramatic rescues in "Bay Watch" style. Such events are fortunately relatively rare. In fact, lifeguards tend to say that a good lifeguard rarely gets wet. Instead most of the lifeguard actions are related to managing the existing risks and to prevent accidents. For instance, in Israel regulations specify that lifeguards may decide to close off certain areas of the beach (post flags) or they may even close the entire designated bathing area for bathing (put up a black flag in the Israeli regulation system). They will do so when entering the water becomes too dangerous, especially due to the presence of rip currents. When the beach is open for bathing, their main activities are "herding", which is directing people away from the rip currents and towards the shore to remain in areas that are relatively safe. These are areas with relatively shallow water, far from rip currents, and relatively close to the shore and to the lifeguard station.

Lifeguards face inherently a multi-objective decision problem. The people who visit the beach would like to have maximal freedom to move in the water. In contrast, the chances of an accident will be minimal if nobody enters the water. Lifeguards need to find a balance between these conflicting goals, taking into account the changing water, wind and wave conditions, the changing population on the beach and the changing viewing conditions. The pay particular attention to the location of rip currents, since these are the major surf hazard in Israel, and they are a major cause of drowning on surf beaches worldwide (Shaw et al., 2014).

Life guards at Tel Aviv beaches

We addressed the study of life guard decision making at a specific location, namely the municipal beaches in Tel Aviv. Tel Aviv has 14 kilometers of shore line on which 13 designated bathing areas have lifeguard stations during the official bathing season from late spring until fall. These stations are staffed according to Israeli regulation by three up to five professional lifeguards who most of the time overlook the beach from a high viewpoint or sometimes from a point on the beach or in the water (usually using a stand-up paddling board). Figure 1 shows a typical view from a lifeguard station. Lifeguards use binoculars to inspect specific events from afar, and they have a powerful directional loudspeaker system with which they can address individuals or groups of people at the surf. Their shifts are typically very long (about 12 hours), and they divide among them the duties during the time at which they are on the job.

Figure 1. View from a life guard station (Ashkelon beach)

Tel Aviv lifeguards face several major challenges. First, the eastern shore of the Mediterranean is particularly unsafe because it is subjected to wave climate enhancing the generation of rip currents, even though the waves are not especially high (Hartmann, 2006 and Hartmann et al., 2009). Israelis, and in particular those who live near the beach, are familiar with these surf hazards, but there are a relatively many drownings of people from higher risk groups, such as tourists and people who live further inland. Second, Israel, and particularly the Tel Aviv area are densely populated, and on some days the beaches can be very crowded. Third, the beaches face west, so in the afternoon and towards the evening the setting sun can make it more difficult to observe events in the water.

To understand the lifeguards' decision making and their considerations regarding risk management we conducted a series of interviews with lifeguards and we had an observer collect observations on 24 days (for about 1 to 2 hours each time) during which he was on the beach and recorded the lifeguard actions, and in particular the times and situation at which a lifeguard decided to intervene. The interviews and observations serve as the basis for our analysis.

Life guard decision making as NDM

NDM has been characterized through four major markers: (1) focus on experienced decision makers, rather than naïve subjects, (2) an array of task and setting factors, (3) focus on situation awareness and not just on the selection of one of the options, and (4) research intended to discover how people make decisions, rather than on the way they should make decisions according to a rational standard (Zsambok, 1997). The first three markers characterize the decision situation, while the fourth has a somewhat different status, as it specifies the purpose of the research and the approach taken.
An analysis of beach life guards' decision making shows that the decisions correspond with the markers. Lifeguards are definitely experienced decision makers, especially on Tel Aviv beaches (where we conducted

observations). Each lifeguard station must have at least two (and will usually have three) lifeguards, and two must have at least 4 years of experience as life guards. Most life guards actually have more than 10 years of experience.

The typical list of task and setting factors characteristic for NDM also fits lifeguard decisions:
1. *Ill-defined goals and ill-structured tasks.* The lifeguard task is highly flexible and ill-defined in that there is often no one specific way a lifeguard should act in a given situation.
2. *Uncertainty, ambiguity, and missing data.* Lifeguards have only limited information about various relevant variables, such as the abilities of individuals on the beach and the exact conditions at different points in the water (because of changing currents etc.).
3. *Shifting and competing goals.* Inherently, life guards must balance the attempt to maximize people's ability to enjoy the beach with the attempt to minimize the possibility of accidents.
4. *Dynamic and continually changing conditions.* Conditions inherently change because people move in and out of the water, move in the water from one place to another, current and wave conditions changes, and viewing conditions change, as for instance, the sun sets and there is glare when looking toward the west.
5. *Action-feedback loops* (real-time reactions to changed conditions). Life guard actions, such as instructing people to move to a certain area, change the conditions in the water and require readjustments on the part of the life guards.
6. *Time stress.* Drowning events develop very rapidly, so interventions to prevent accidents need to be done quickly and as immediate responses to developments in the water.
7. *High stakes.* Drowning accidents can be fatal, and even if they are not, they can cause severe permanent injuries.
8. *Multiple players.* The actions of a lifeguard are inherently social, interacting with the people on the beach, taking into account the possible implications of accidents in the eyes of supervisors or legal authorities.
9. *Organizational goals and norms.* The organizational goals are competing, furthering a culture of beach life while at the same time stressing safety.

An important aspect of the lifeguard's work is the management of situation awareness. Lifeguards must actively regulate observations, for instance by dividing areas of observation between different guards. They may also move outside the station to get a better view of certain areas, at times using stand-up paddle boards to be closer to possible danger spots.

While lifeguard decision making shares many characteristics of a typical NDM problem, it is also special in some respects. First, the actions are routinely done for hours on a regular basis. This is different from most NDM problems (such as firefighting, military operations, surgery, etc.), where the decisions are usually made in a unique, often extreme situation. Here the monitoring and intervening is done at a specific, constant location over prolonged periods of time.

Also, and relatedly, the situations and events tend to be fairly predictable. Although there is an infinitely large set of specific events, basic patterns will often repeat themselves, allowing the lifeguards to use simple decision rules. Thus, here, perhaps more than in domains where events are highly unique, a recognition-primed decision technique (Klein, 1998) is likely to be possible.

OUTLINE OF A POSSIBLE MODEL

The nature of a lifeguard's task makes it relatively easy to develop a normative model. We present here the outline of such a model without developing the mathematical expressions. The normative model of life guard decision making must address the different variables the life guard should consider when deciding on an action. These variables can be broadly divided into three groups of factors:

1. *Environmental factors.* These include beach characteristics, weather and wind conditions, wave heights and frequencies, existence of currents, and overall viewing conditions. Environmental conditions can range from least hazardous to extremely hazardour, according, for instance, the Beach Hazard Rating index (BHR; Short & Hogan, 1995) or the Onshore Storminess Factor (ONSF; Hartman, Pick & Segal, 2009).
2. *Beach user factors.* These are factors characterizing the population of beach users at a given moment. They include the number of people (which changes with daily, weekly and seasonal fluctuations), the presence of specific risk groups, such as children, tourists, people under the influence of alcohol or drugs, the level of swimming skill people have, as well as the knowledge people have about beach hazards.
3. *Lifeguard factors.* These are factors related to the lifeguards themselves, their number and positions, the equipment available at a given time, etc.

These three factors can be seen as three dimensions of a state-space. Each axis ranges from low risk to high risk. See Figure 2 for a depiction of this space.

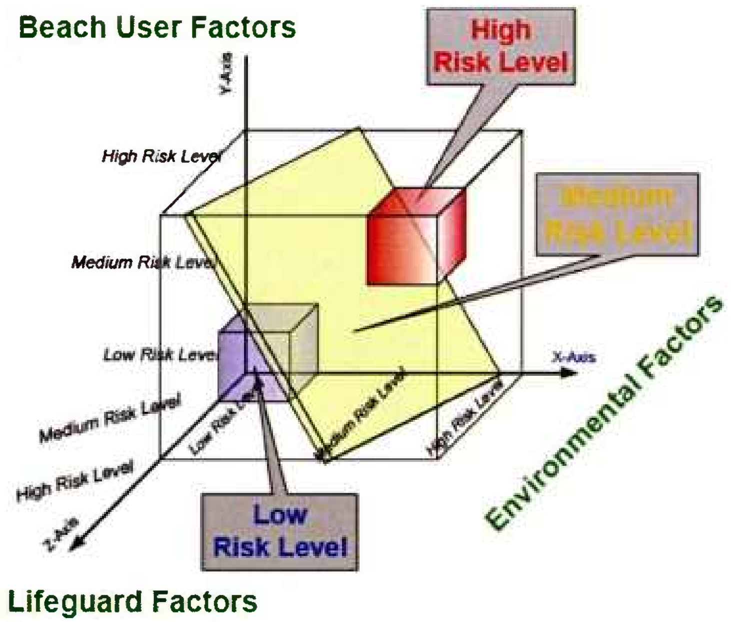


Figure 2. Three-dimensional risk space.

At any moment in time, with particular beach users, the system is at a specific position in this space. Some positions, for instance when all three dimensions have high risk values (e.g., high waves, inexperienced beach users, only few lifeguards) are associated with high risks. Other position can have lower or even very low risk (as when the sea is calm and there are only few and experienced beach users on the beach). At each position, there is a value of risk associated with the position, as well as a value of the benefit a person obtains when being at the position.

It should be noted that the position is not a single value in the three-dimensional space for a given moment in time, but rather it is a set of values, one for each beach user or group of users. So, for instance, at a given moment a family with small children will be at a different point in the space than a nearby experienced adult swimmer.

The lifeguard in this model evaluates the position of each point, relative to some decision criterion. In its simplest form, the criterion is a plane in the three dimensional space which divides the space into a region of high risk and a region of low risk. If a particular event is seen as being at a position that is riskier than the threshold risk, the lifeguard can intervene and change its position. This can be done by the lifeguard actively moving closer to the person, thereby making it easier to intervene if a danger situation should develop. More frequently, however, the lifeguard will change the position of the beach users, asking them to move closer to the beach, towards another area in the water or out of a danger zone.
In its simplest form, the lifeguard evaluates each point independently, without considering the existence of other points. One of the challenges lifeguards face is the need to make a distinction when and how to intervene when different user groups share the beach. Limiting all activities to the optimal level for the least able user group will greatly lower the benefits others may have from being at the beach. However, allowing some people to do something (for instance, going relatively deep into the water) may signal to others that it is safe to do so, and this may cause a problem. Thus, the different points are not independent, and an expanded optimal model may have to take such interdependencies into account.

CONCLUSIONS

We present here the specific case of lifeguard decision making as an example of naturalistic decision making. It is a somewhat atypical example, as it involves frequent repeated decisions over long periods of time. We show that it is possible to define a quantitative model of the decision that can be used to compute the optimal response strategies, given specific risk conditions and values.

We do not claim that lifeguards actually make such computations. This is highly unlikely, as solving such problems is computationally complex. Thus the multi-objective decision model is not supposed to be a descriptive, psychological model, describing the cognitive processes involved in the decision making. Rather, it is a normative model that can generate predictions to which one can compare the actual decisions. The existence of such differences does not necessarily mean that lifeguards function incorrectly, but they may raise awareness for possible points in which decision making can be improved. This is an important issue, so far often absent from NDM research, where decision training often means moving novice decisions closer to expert decisions.

This is based on the assumption that the expert decisions are necessarily correct and are not subject to systematic biases.
The possible use of quantitative models that capture situations usually addressed withing the framework of NDM is not limited to lifeguard decision making. It may actually be a general challenge for research on decision making. NDM is not an antithesis to quantitative models. Rather, it shows problems that differ from those normative decision models can solve relatively easily (e.g., choices among gambles). These are problems modeling researchers will have to address if they want their work to be relevant for the complex decision situations that are the subject of NDM research.

ACKNOWLEDGMENTS
We thank Assaf David, Yotam Green and Arnon Malka for their help in earlier work on this topic.

REFERENCES

Hartman, D. (2006). Drowning and Beach-Safety Management (BSM) along the Israeli Mediterranean beaches of Israel – a long-term perspective. *Journal of Coastal Research*, 22, 1505–1514.

Hartmann, D., Pick, K. & Segal, Y. (2009). Onshore Storminess Factor: A new tool for regional beach hazard rating and beach safety management. *Journal of Coastal Research*, SI 56, 807-811.

Klein, G. (2008). Naturalistic decision making. *Human Factors*, 50(3), 456-460.

Klein, G. A. (1998). *Sources of Power: How People Make Decisions*. MIT Press: Cambridge, M.

Shaw, W. S., Goff, J., Brander, R., Walton, T., Roberts, A., & Sherker, S. (2014). Surviving the surf zone: Towards more integrated rip current geographies. *Applied Geography*, *54*, 54-62.

Short, A. D., & Hogan, C. L. (1995). Rips and beach hazards, their impact on public safety and implications for coastal management. *Journal of Coastal Research*, 12, 197-209.

Szpilman, D., Bierens, J., Handley, A., & Orlowski, J. (2012). Drowning. *New England Journal of Medicine*, 366, 2102-2110.

Zsambok, (1997). Naturalistic decision making: Where are we now? In C. E. Zsambok & G. Klein, *Naturalistic Decision Making*. (Chapter 1). Lawrence Erlbaum Associates.

Expertise Management: Challenges for adopting Naturalistic Decision Making as a knowledge management paradigm

Brian M. MOON[a], Holly C. BAXTER[b], and Gary KLEIN[b]
[a]*Perigean Technologies LLC*
[b]*Strategic Knowledge Solutions, MacroCognition LLC*

ABSTRACT
Naturalistic Decision Making (NDM) offers perspectives and methods for understanding cognitive performance, especially expertise. Knowledge management (KM) is a multi-disciplinary field professing to improve organizational performance by making the best use of knowledge. Since expertise is one of the most valuable assets in any organization, it stands to reason that adopting the perspectives and methods of NDM as a KM paradigm – i.e., Expertise Management – should enable organizations to realize performance improvements. Yet, Expertise Management has not achieved recognition as a KM strategy, and attempts to implement it have been met with significant methodological, practical, and competitive challenges. This paper examines the case for NDM-based Expertise Management as a core KM strategy, and the methodological, practical and competitive challenges for adoption. The authors draw on their collective professional experience in attempting to implement Expertise Management at a diverse range of organizations, and conclude with recommendations for future directions.

KEYWORDS
Expertise; Expertise Management; Macrocognition; Cognitive Task Analysis; Business; Practical Application.

INTRODUCTION

Naturalistic Decision Making (NDM) offers perspectives and methods for understanding cognitive performance, especially expertise. Perspectives include a focus on macrocognition (Klein et al., 2004), while methods include cognitive task analysis (Crandall, Klein and Hoffman, 2006; Moon et al., 2011). NDM has demonstrated value for revealing the nature of expertise in diverse domains, and providing guidance for translating expertise into the design of systems for enhancing individual and team performance (Klein, 2008).
Knowledge management (KM) is a multi-disciplinary field professing to improve organizational performance by making the best use of knowledge. Duhon (1998) offered one of many comprehensive definitions of the field:

> Knowledge management is a discipline that promotes an integrated approach to identifying, capturing, evaluating, retrieving, and sharing all of an enterprise's information assets. These assets may include databases, documents, policies, procedures, and previously un-captured expertise and experience in individual workers.

The last point expresses relevance for NDM. Since expertise is one of the most valuable assets in any organization, it stands to reason that adopting the perspectives and methods of NDM as a KM strategy should enable organizations to realize the performance improvements at which KM is targeted. The notion of an NDM-based practice of KM was first suggested by Klein (1992), observing that while "(e)xpertise is a key resource in any organization…it is usually not treated with the same care as other resources" (p. 170). Klein described the very situation that KM purports to address: "Few organizations have any methods for preserving or expanding their experience, or even taking stock of their current expertise…When staff members retire, the organization does little to preserve their expertise (p. 170)." The latter insight has become particularly important as the world's workforce skews toward massive retirement (Hoffman & Hanes, 2003). Klein's primary recommendation was the use of "low-technology applications of knowledge engineering" (p. 170), or knowledge elicitation. Specifically, Klein advocated using the Critical Decision Method to capture incident accounts and lessons learned (p. 180-184), building a case that "the various methods of knowledge engineering have the potential for maintaining corporate memory and for preserving organizational expertise (p. 185).

Yet, NDM has not achieved wide recognition as a KM strategy. We can take Koenig's article, "What is KM? Knowledge Management Explained" (2012), published for KMWorld, as an exemplar of the many reviews of the field. It is notable that despite expertise being explicitly cited as a core concept in Duhon's definition, the nature of expertise is only tangentially referenced through Nonaka & Takeuchi's (1995) oft-cited distinction between explicit, implicit and tacit knowledge. Moreover, of the three undertakings that Koenig suggests are "quintessentially KM," none offers guidance on the articulation or acceleration of expertise. Lessons learned databases, expertise location, and communities of practice are suggested as mechanisms to enable the capture, location and sharing of expertise – in much the way other, more tangible corporate assets are handled. Koenig's historical view of the stages of development of KM is also illustrative: first stage—information technology; second stage—HR and corporate culture; third stage—taxonomy and content management. Noticeably absent is reference to any affiliation with the traditions of NDM, expertise studies or even cognitive engineering and human factors.

Several reasons underlie the apparent lack of recognition of the potential value of NDM for KM. The major figures in the field of KM derive from fields such as business studies, organizational theory, management consulting, education, library sciences, and information technology. While these leaders often express a familiarity with elements of the NDM paradigm – e.g., "collective sensemaking" (Dixon, 2015) and knowledge capture – they rarely cite NDM research as the basis of their understanding of expertise. Ackerman and Wulf et al., (2003), while concerned with "sharing expertise," rely principally on judgement and decision making paradigms to guide their views on expertise and computer-supported cooperative work to formulate approaches for getting "beyond knowledge management." There are exceptions as some have recently realized the potential advantages of knowledge elicitation techniques (Gavrilova and Andreeva, 2012). Another reason is the significant influence of technology-based solutions providers in the KM community. The first stage of KM continues to command significant focus from organizations seeking help with KM challenges; thus, any KM approach that is principally human-centered often struggles to gain audience (Griffiths and Moon, 2011).

But perhaps the primary reason that NDM has not emerged as a core KM strategy is because attempts to implement it have been met with significant methodological and practical challenges. These challenges are the focus of this paper. Collectively, we have gained professional experience in attempting to implement an NDM-based KM strategy – i.e., Expertise Management (EM) – at a diverse range of organizations. Our experience includes:

- Multi-year and pilot implementations, and training events, for major corporations in the energy and manufacturing industries and government sector,
- Publications and methodological guidance documents,
- Faculty appointment and curriculum development for a graduate-level course in EM and Knowledge Elicitation, and
- Experimentation with a ShadowBox method for capturing and disseminating expertise.

Our experience has revealed the challenges, and propelled us to refine our Expertise Management approaches. We conclude with recommendations for future applied research directions.

METHODOLOGICAL CHALLENGES

To contextualize our review of the methodological challenges, it is useful to present a general model of EM, as we have implemented and taught it. Our model comprises three elements: identify, articulate and engage. Identify refers to approaches for identifying expertise on which to focus the subsequent activities. Articulate references knowledge elicitation activities, typically conducted by an experienced knowledge elicitors working one-on-one with identified experts. And Engage covers activities that are intended to facilitate the acceleration of expertise in others, to include sharing the articulated expertise through representations and training exercises. In some cases, the identification of expertise is straightforward – target experts have already been identified by the sponsor or manager of the effort. In other case, organizations have needed a principled approach to scaling proficiency in order to focus the effort (see Hoffman, et al., 2014). The bulk of our efforts have fallen under the articulate element, where we have either conducted or trained others how to conduct knowledge elicitation. The engage element has at times either been beyond the scope of our efforts, or pre-determined by the organizations with whom we have worked. In some cases, we have also blended the articulate and engage activities by bringing others into the articulation process so that engagement is concurrent (Baxter, 2013). Recent efforts have seen us focus explicitly on the efficacy of some of our engagement activities such as the ShadowBox method (Klein, Hintze & Saab, 2013).

The methodological challenges to NDM as a KM strategy fall into three categories: scope, process, and product.

Scope

By scope, we mean the challenge we often hear from experts and business leaders at the start of our EM engagements: "How are you possibly going to capture everything I've learned in 30 years?" The question is a reasonable one, coming from professionals who have achieved "franchise expert" status (Hoffman et al., 2011), earned through years of compiled experience – yet have very little insight into knowledge elicitation or the purpose of the EM effort. Indeed, while managers may realize the risks of "lost knowledge" (DeLong, 2004) and want to take steps to mitigate it, they too are often not clear about what they can do or even what they want to achieve. Yet they often fail to define what specifically about the expertise is of interest and what will be done with the knowledge after it is captured. We have learned this lesson the hard way, for example after being introduced to nuclear engineers with experience in "instrumentation and controls" or "fuels" (Moon and Kelley, 2010) – huge practice areas implicating vast subdomains, skillsets and tasks requirements. It is very difficult to know from the outset where the most critical macrocognitive elements of performance may lie.

The NDM paradigm is mostly silent on the issue of scoping an EM effort. NDM provides a methodological toolkit for understanding the expertise inherent in the proficient performance of tasks, which can be extrapolated to the understanding of roles (Crandall, Klein and Hoffman, 2006). But determining which aspects of an expert's experience hold the most potential for realizing a return on the investment of resources in an EM engagements requires honing in on the expert's career, the current and envisioned needs of the organization, and perhaps most importantly the needs of the personnel that will take up the expert's responsibilities. Hoffman and Hanes (2003) advocated for a process that focuses on the elicitation of knowledge that is (a) unique to the individual expert, (2) crucial for the organization and (3) not currently documented—yet even this approach can yield significant candidate topics. EM efforts will quickly and invariably expand from a focus on the individual expert to the broader context of the organization.

Scoping is also a challenge at the other end of the EM engagement; that is, knowing when to stop. While many CTA interviews typically last something in the range of two hours for a single task or incident, we know from experience that collection and analysis of some protocols and case studies can take many hours (c.f., Hoffman et al., 2000). Experienced knowledge elicitors have heuristics to inform them about when an incident or topic has been thoroughly covered. But when engaged in an EM effort, there are no analogous rules of thumb for how long an EM engagement should last, or when it is "done." There is always another incident that could be captured; always another aspect of the expert's experience that could be explored. We have been fortunate in some engagements to spend upwards of 30 to 40 hours over the course of six to eight months with some experts (Moon and Kelley, 2010) – a luxury in the study of expertise. More often than not, EM engagements end much earlier for practical reasons, not the least of which is the always present need to get the expert back to work. Resources also impose constraints on the scope.

Process

The general model of EM that we sketched above could be expanded to show a number of subprocesses that we have managed in our EM engagements. Two in particular have shown to be methodologically thorny. The first regards building and maintaining rapport with the expert. Generally speaking, many of the experts we have worked with are favorably disposed to the idea of preserving some of their critical knowledge and helping others gain some advantage from it. They are motivated by senses of personal and professional legacy and a desire to see the organization succeed. We have, however, been met with experts who were quite unmotivated to participate. In one striking case during what was supposed to be a pilot demonstration, an expert approached one of the authors on the eve prior to the pilot and stated bluntly, "I don't want to do this – I think it is a bunch of crap." He later revealed that he felt strongly that his junior colleagues should learn their science and craft the same way he did – through "hard work and getting their hands dirty." This expert's lack of motivation had significant methodological implication on the EM effort, as he was mostly unwilling to engage with any of the knowledge elicitation methods that were to be demonstrated.

A second challenge has been the selection and execution of knowledge elicitation methods during the articulation activities. The Critical Decision Method (Hoffman, Crandall and Shadbolt, 1998) and Applied Concept Mapping (Moon et al., 2011) have been our KE methods of choice because of their established track records in capturing macrocognitive elements of performance. They have not, however, always worked well for our purposes. CDM has been challenging to apply with experts whose key value to the company lied in their vast declarative knowledge – e.g., about historical customer and vendor relationships – and with experts whose tactical experience was so vast that they were challenged to recall any particular incidents. The latter point is particularly important in light of the scoping issue. The initial CDM question, "Can you think of a time when your skills were challenged?" has quite often garnered a response of "many, many times," putting the onus back on the knowledge elicitor to help the expert scope his or her recall. We have learned that starting an EM engagement with a CDM interview is not an efficient way to scope the effort.

Applied Concept Mapping introduces a number of process complexities that have been discussed in detail elsewhere (c.f., Moon, Hoffman, Eskridge and Coffey, 2011). At times, these complexities have overridden the potential value of use. For one example, one of the authors was engaged for EM with a nuclear engineer whose career dated back to the dawn of the nuclear age. The engineer was born and raised in Japan, and the legacy of his first language remained quite evident in his accented English. At an age when many workers would be well into their second decade of retirement, this expert reported for assignment five days a week, adding to his already prolific research and publication achievements – preferring to work mostly behind the closed door of his office. While concept maps articulating aspects of his vast declarative knowledge and the reasoning strategies that helped shape his industry-altering ideas would almost certainly have created value for the organization, executing the protocols for concept mapping would have been very difficult with this expert.

There exists a tacit assumption in the NDM paradigm with regard to methods for understanding expertise: namely, that the methods can be executed with any expert, under any conditions. While caution for the assumption has been given by suggesting adaptation will always be necessary, our attempts to implement EM have revealed several boundary conditions that have served to bring the assumption into high relief.

Products

A goal of EM is to externalize expertise so that it can be preserved in ways that enable others to engage with it. The traditional representational and analysis products suggested by the NDM community include decision ladders, decision requirements tables, concept maps, and timelines (c.f., Crandall et al., 2006). While these products have proven useful for NDM practitioners in order to guide design and development activities, their value for EM has been difficult to demonstrate. In the context of EM, we have prepared products ranging from knowledge models, which are hyperlinked sets of concept maps and associated knowledge resources (Hoffman and Beach, 2013), to extensive incident accounts, to narrative content formatted by client requirements for corporate intranets and other KM portals. Moreso than is typical of our other NDM work, our EM products have often benefited from iterative review with our experts, though such reviews have introduced additional methodological and practical concerns such as how to reign in an expert's revisions (Moon and Kelley, 2010).

Applied Concept Mapping has presented particular challenges. Very little study has been made of the efficacy of concept maps for helping accelerate the achievement of expertise. Moon and Hoffman (2008) demonstrated the potential value for concept maps for "rapid idea transfer" in military populations showing slight improvement over PowerPoint but lower efficacy compared to narrative. Yet Derbentseva and Kwantes (2014) have only seen a "lukewarm response to using Cmaps for communicating information" in the same population. There are also practical concerns. Coffey and Eskridge (2008) have noted the "format problem" with concept maps, particularly in the industries where the "format of training materials and procedures is clearly circumscribed and tightly controlled" (p. 17).

While NDM points to the nature expertise, to techniques for its analysis and representation, and to methods accelerating its achievement (Hoffman et al., 2014), guidance for how to preserve and present externalized macrocognition in ways that permit efficient, flexible, context-situated exploration as a means toward acceleration has been underspecified.

PRACTICAL CHALLENGES

Some practical challenges were alluded to above, e.g., time, resources, and regulations. These are not new to NDM-based efforts. Our experiences implementing EM have forced us to confront several other practical challenges dealing with making the case for it and the need for NDM expertise.

Making the case

Making the case for an NDM-based EM is challenging for several reasons. Solving or mitigating problems through an NDM approach is, almost by definition, a time consuming and thus expensive endeavour (CITE). Expenses are introduced through the expert's and EM expert's time, as well as travel and other expected costs. Personnel charged with mitigating lost knowledge must typically develop a cost/benefit analysis in order to answer the return on investment question. The analysis may take other KM "solutions" into account, including software products, mentoring programs and succession plans – none of which take seriously the need to deeply understand expertise. While some decision makers in human resources and training departments may be familiar with NDM traditions and requirements for success, the vernacular of NDM (e.g., "macrocognition") does not translate well to front-line managers who stand to gain the most from it. Indeed, as noted above, even many KM practitioners do not speak the language.

The case has also not benefitted from the success stories that have emerged from other applications of NDM, such as cognitive engineering (c.f., Cooke and Durso, 2010). EM is at a stage of development just behind "accelerated learning" (Hoffman et al., 2014). We have many examples of application, including institutionalization at some organizations (c.f., Kelley, Sass and Moon, 2013). We know of many anecdotal examples of success, where "management gained real insight into what the experts truly did, and in many cases, much greater insight into how they did it than they could have had without [EM]" (Coffey and Eskridge, 2008). What remains missing is systematic analysis of the effects and benefits of EM, and measurement of the costs and risks of *not* doing EM.

Need for NDM expertise

Our implementation efforts have included educating KM students and training and coaching personnel who have taken on KM as primary and collateral duties. In more than a few cases, we have introduced EM to personnel who inherited KM yet had no prior experience in KM or exposure to EM. We have found that some people take quickly to some methods but struggle with others. Some have seen uses for the methods in their other work, beyond EM. During one engagement, one of the authors was coaching two candidate EM practitioners who inherited KM as a collateral duty. After several days of coaching in CDM and Applied Concept Mapping, it became quite clear that one of the candidates was quite skilled at formulating and asking questions while the other was very skilled at creating and editing concept maps. Yet neither was very good at the other skill – to the point of bringing interviews to a halt in order to shift roles back to those they were comfortable in.

Many of our trainees have expressed appreciation of the value of the insights into expertise. They have seen first hand through demonstration of knowledge elicitation just how deeply the techniques can unpack expertise. Once they see an interview unfold, they are often surprised at the fluidity with which a skilled elicitor can exercise methods in order to unveil details that would have otherwise not been articulated.

Yet our training and coaching experience has driven home the point that in order to effectively and efficiently work with experts who will be the concern of an EM program, practitioners need a deep level of familiarity with NDM and a flexible facility with its attendant toolkit. This need for NDM expertise introduces a significant practical challenge to the proliferation of NDM as a KM paradigm. Indeed, the scope of the expertise loss problem is many orders of magnitude larger than the entire NDM community could service. For KM practitioners who are not steeped in NDM, we have seen the struggle to get up to speed in NDM quickly enough to be effective. Given that there are not many NDM practitioners who offer their work as motivated by KM goals, reaching the tipping point of adoption will require new directions.

FUTURE DIRECTIONS

We are encouraged about the prospects for NDM adoption as a KM paradigm by the immense opportunities that KM challenges present. As the global population continues to age, the need for at the very least preservation of expertise will only become more critical. Stories of some organizations losing expertise – e.g., NASA's ability to fly Saturn 5 rockets (DeLong, p. 12) – will continue to amass. Demand will drive the search for KM approaches that show promise for mitigating the problem.

We are also encouraged by the successes of our limited efforts, where success can be measured in client value and empirical findings. Notably, we have realized several of our success stories through adaptations and extensions from our general model. Kelley, Sass and Moon (2013) reported on the maturation and modification of our general model that outlined four types of elicitation sessions: technical content of interest to many, job replacement, facilitation, and critical skill transfer. Each type introduced a different purpose, resource requirements and audience, but every type utilized knowledge elicitation techniques. Client value was realized in "high levels of expert engagement… consistent recommend[ation of] the process to their peers and management" (p. 4). The approach was "viewed as a valid alternative to one-on-one mentoring" (p. 4). In other efforts, we have adapted the model into one-off sessions that are scoped by the client. Value for these clients has been evident through the long-term engagements they have established with us—clearly, these organizations are realizing value, even if organizational effects have not been systematically measured.

Future directions will need to more systematically compare and contrast implementation models of EM and their relative values. Independent variables for exploration should include time spent with experts, knowledge elicitation techniques, and experience of the knowledge elicitors. Dependent variables should focus on realized improvements, preferably at the organizational level. McManus, Wilson and Snyder (2003) demonstrated positive revenue benefits from a "knowledge harvesting" approach, showing that bottom line improvements are possible. We would also expect to see improvements in macrocognitive performance in organizations. We are very encouraged by recent efforts toward such systematic investigation. Klein et al. (2015) found strong effects using ShadowBox scenarios to help military personnel acquire tacit knowledge in the form of a new sensemaking frame. ShadowBox addresses some of the product challenges, and we are eager to explore how the approach can scale for organizations in the Engage stage.

Expertise Management offers potential to improve organizational performance through the introduction of the NDM paradigm into KM programs. We have accumulated lessons learned from the many challenges to implementation. It is our hope that these lessons and future applied research will help establish EM as a viable KM—and NDM—practice area.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the sponsors of the work referenced herein. We also wish to thank Robert Hoffman for his thoughtful review of early drafts, and for his leadership and partnership in EM.

REFERENCES

Ackerman, M. S., Pipek, V., & Wulf, V. (Eds.). (2003). *Sharing expertise: Beyond knowledge management.* MIT press.

Baxter, H. C. (2013). Transferring Specialized Knowledge: Accelerating the Expertise Development Cycle. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)* (Vol. 2013, No. 1). National Training Systems Association.

Coffey, J. W., & Eskridge, T. (2008). Case studies of knowledge modeling for knowledge preservation and sharing in the US nuclear power industry. *Journal of Information & Knowledge Management*, *7*(03), 173-185.

Crandall, B., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis*. Mit Press.

DeLong, D. W. (2004). *Lost knowledge: Confronting the threat of an aging workforce*. Oxford University Press.

Duhon, B. (1998). It's all in our heads. *Inform*, *12*(8), 8-13.

Gavrilova, T., & Andreeva, T. (2012). Knowledge elicitation techniques in a knowledge management context. *Journal of Knowledge Management*, *16*(4), 523-537.

Griffiths, D., & Moon, B. (2011). The State of Knowledge Management. *KM World Magazine*. Volume 20, Issue 10.

Hoffman, R. R., Ward, P., Feltovich, P. J., DiBello, L., Fiore, S. M., & Andrews, D. H. (2014). *Accelerated Learning: Training for High Proficiency in a Complex World.* Psychology Press.

Hoffman, R. & Beach, J. (2013). Lessons learned across a decade of knowledge modeling. *Journal for Educators, Teachers and Trainers*, Vol. 4 (1), pp. 85 – 95.

Hoffman, R., Ziebell, D., Feltovich, P., Moon, B., & Fiore, S. (2011). Franchise Experts. IEEE Intelligent Systems, Vol. 26, No. 5, pp. 72-77.

Hoffman, R.R. and Hanes, L.F. (July-August 2003). The boiled frog problem. *IEEE Intelligent Systems,* pp. 68-71.

Hoffman, R. R., Coffey, J. W., & Ford, K. M. (2000). A case study in the research paradigm of human-centered computing: Local expertise in weather forecasting. *Unpublished Technical Report, National Imagery and Mapping Agency.*

Hoffman, R. R., Crandall, B., & Shadbolt, N. (1998). Use of the critical decision method to elicit expert knowledge: A case study in the methodology of cognitive task analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *40*(2), 254-276.

Kelley, M., Sass, M., and Moon, B. (2013). Maturity of a Nuclear-Related Knowledge Management Solution. Annual Meeting of the American Nuclear Society. June 16-20.

Klein, G. (1992). Using knowledge elicitation to preserve corporate memory. In R. R. Hoffman(Ed.), *The psychology of expertise: Cognitive research and empirical AI* (pp. 170–190).Mahwah, NJ: Erlbaum.

Klein, G. (2008). "Naturalistic Decision Making". *Human Factors: The Journal of the Human Factors and Ergonomics Society* 50 (3): 456–460.

Klein, G., Borders, J., Wright, C., & Newsome, E. (2015). An empirical evaluation of the ShadowBox training method. Paper submitted for presentation.

Klein, G., Hintze, N., & Saab, D. (2013). Thinking inside the box: The shadowbox method for cognitive skill development. Proceedings of the 11th International Conference on Naturalistic Decision Making (NDM-11). Marseille, France.

Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (2003). Macrocognition. *Intelligent Systems, IEEE*, *18*(3), 81-85.

Koenig, M. E. (2012). What is KM? Knowledge management explained. *Retrieved November*, *6*, 2013.

McManus, D. J., Wilson, L. T., & Snyder, C. A. (2003). Assessing the business value of knowledge retention projects: results of four case studies.

Moon, B., Hoffman, R. R., Novak, J., & Canas, A. (Eds.). (2011). *Applied concept mapping: Capturing, analyzing, and organizing knowledge*. CRC Press.

Moon, B., Hoffman, R., Eskridge, T. & Coffey, J. (2011). Skills in Concept Mapping. In Moon, B., Hoffman, R., Novak, J. and Cañas, A. (eds.) *Applied Concept Mapping: Capturing, Analyzing, and Organizing Knowledge*. pp. 23-46. New York: CRC Press.

Moon, B., & Kelley, M. (2010). Lessons Learned in Knowledge Elicitation with Nuclear Experts. Invited Paper. Seventh American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies. Las Vegas. Nevada. American Nuclear Society, LaGrange Park, IL. November 7-11.

Moon, B. & Hoffman, R., (2008). Rapid and Accurate Idea Transfer: Presenting Ideas with Concept Maps. Scientific and Technical Final Report. Contract No. W31P4Q-08-C-0229. Washington DC: Defense Technical Information Center.

Nonaka, I. & Takeuchi, H. (1995). *The knowledge creating company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press.

Exploring Cue Use in Rugby League Playmakers to Inform Training Initiatives

David JOHNSTON[a], Ben MORRISON[ab]
[a] *Australian College of Applied Psychology*
[b] *Macquarie University*
E: david.a.johnston2@gmail.com
P: +61 4 23198327

ABSTRACT
Within the sport of Rugby League there exists a perceived shortage of talent in playmaking positions (i.e. halfback, five-eighth and hooker) (Barton, 2013). An academy dedicated to the development of playmaking skills has recently been established (Proszenko, 2013). The precise skills targeted by the academy for development have yet to be determined. The purpose of this research was to investigate the nature of cue use in decision making in a Rugby League context and determine whether players of differing ability could be differentiated to inform potential cue-based training initiatives. Rugby League playmakers were interviewed using a variation of the Cognitive Task Analysis, which employed a picture stimulus. The sample consisted of 10 players, six of whom played with a professional Rugby League Team and four from an amateur Rugby League Team. Directed content analysis was performed on the resulting transcripts, and concept maps, cognitive demands tables and a critical cue inventory was produced. Results indicated that professional players demonstrated greater cue discrimination, assigned different meaning to the cues and processed cues in a different manner to amateur players. The results offer insights for future design of training programs in the development of playmaking skills, and raise important questions regarding the use of critical cue inventories in training.

KEYWORDS
Research/Experimentation; Cognitive Task Analysis; Decision Making; Sport

INTRODUCTION
Within the sport of Rugby League there exists a perceived shortage of skilled playmakers (i.e., halfback, five-eighth and hooker) (Barton, 2013). Consequently, an academy, dedicated to the development of playmaking skills has recently been established (Proszenko, 2013). However, the precise skills targeted by the academy for development are yet to be determined.

It could be argued that decision-making, how people select one option from a set of possible options(Tenenbaum & Bar-Eli, 1993), is a critical skill for a Rugby League playmaker. Specifically, being able to estimate or predict some aspect of the environment on the basis of available cues would seem intuitively important to such a rapidly played out sport. Cues constitute a relationship held in memory between environmental features and/or events that hold some meaning or value to the individual (Ratcliff & McKoon, 1994).

It has been hypothesised that cue selection/use results from previous environmental experience (Wiggins, 2006). The value of using appropriate cues in sports has been shown in temporal occlusion studies, which show that when particular body segments were occluded, there is a significant decrease in prediction accuracy (Shim, Carlton, Chow, & Chae, 2005). Thus, it would appear that a characteristic of superior decision-making performance may be reflected in an individual's capacity to identify prospective features, ascribe meaning to salient events and avoid being distracted by other features (Shanteau, 1992).

The importance of cues to performance has been highlighted in research demonstrating that superior performers can be differentiated by the number of cues used (Shanteau & Hall, 1992). When both relevant and irrelevant information cues are given, experts are better at cue discrimination, only selecting and using the relevant cues (Shanteau & Hall, 1992). It has been suggested that too many cues may complicate an array of information, impinging upon limited information processing resources. Cue discrimination may therefore allow for superior performance. For instance, research which examined batting in cricket has revealed that highly skilled cricket players use cues from the bowler to assist in decision making (Renshaw & Fairweather, 2000). Interestingly for highly skilled batters, it is not cues in isolation which provide the most information rather the cues in association to one another which leads to more accurate predictions (Renshaw & Fairweather, 2000). This cue clustering has been observed in expert criminal investigators (Morrison, Wiggins, Bond, & Tyler, 2013) and has been suggested as being linked to formations of highly developed domain-specific memory structures (Yarrow, Brown, & Krakauer, 2009).

There has also been observed differences in the manner that cues are processed by expert and novice decision makers. Greitzer, Podmore, Robinson, and Ey (2010) observed significant differences between novice and expert power system operators in how cues were processed. The authors observed that novices would respond to cues and patterns at a rule based level (Greitzer et al., 2010). Novices reacted to disturbances in an effortful conscious manner consistent with applying a pre-packaged unit such as an "If X cue then Y response"(Greitzer et al., 2010). This was considerably different to the expert operators who processed information at the skill-based level, reacting to these cues at an automatic subconscious level (Greitzer et al., 2010). Cues were used preventatively rather than reactively and the behaviour was executed with little conscious thought (Greitzer et al., 2010).

Within the sporting literature, research has suggested that players control a situation by focusing on salient cues which allow them to make the most appropriate decision (Macquet & Fleurance, 2007)Within the sport of Rugby League a further study has suggested that cues may play an important role (Gabbett & Abernethy, 2013)A study compared elite and semi-elite players in their ability to react to Rugby League defensive scenarios which were projected onto a screen (Gabbett & Abernethy, 2013). Participants were compared with respect to their response times and the accuracy with which they responded to the stimulus. The study found that highly skilled players had faster response times and had greater accuracy in their responses (Gabbett & Abernethy, 2013). It was suggested that this demonstrated the ability of experts to recognise relevant game-specific cues to which the lesser-skilled players were not attuned (Gabbett & Abernethy, 2013). To date no research has examined what cues allow for this improvement in performance. Similarly, the nature of the study did not assess whether differences existed in how the cues were used as a function of skill level, nor did it explore how such fast assessments are made possible within a decision making system.

The present study aimed to examine the extent to which cues are used within the sport of Rugby League. From the research outlined, it was anticipated that cues would be used as a means to assist in decision making. The research also aimed to determine whether cue use changed as a function of the level of expertise. From the research examined it was anticipated experts would practice greater cue discrimination than novices. It was also hypothesised that experts would process, assign meaning and use these cues differently to novices.

METHOD

Participants

The participants consisted of players from a semi-professional rugby league club (N=3) and players from a professional *National Rugby League* Club (N=7). The representative structure which players progress through represents a continuum with which to judge relative expertise in players. This system discriminates based on ability and only those who demonstrate consistent replicable superior performance progress to a higher grade. Although differences between grades have not been quantified explicitly they do represent an objective measure of different competencies and were used to categorise players. Players were assigned categories based on their grade, with category 1 representing the highest level of ability and category 4, the lowest. See Table 1. for distribution of participants across gradings and their assigned player category.

Table 1. Distribution of Participants in Study

Team	Grade	Number of Participants	Player Category
Professional	First Grade	1	1
	Reserve Grade	3	2
	National Youth Competition	3	3
Semi-Professional	First Grade	1	4
	Second Grade	2	4

A semi-structured interview was conducted based on Cognitive Task Analysis (CTA) methodology. CTA methods were used to elicit the cues and contextual considerations influencing judgements and decisions (Militello & Hutton, 1998). The researchers used probes outlined by Militello and Hutton (1998), to help identify a range of cues and patterns. Questions were modified to align with the sport of Rugby League and the modifications were piloted on a Rugby League player. Prior to the commencement of the interviews, the modifications were verified as effective by a trained CTA expert.

Picture Stimulus

A picture stimulus was used in conjunction with the CTA as a means to stimulate further discussion and increase understanding of problem solving methods (Morrison et al., 2013). CTA methods rely on recall of an event in the player's history, which may limit the ability of researchers to compare between different participants. The stimulus created a unique opportunity to compare players of differing ability across a common scenario as well as corroborate knowledge which was elicited from the CTA interview.

The picture stimulus (Figure 1) was a scene from a previous played First Grade professional rugby league match. The scene was selected due to the lack of structure in the defence which meant that there were many available options. The problem had a 'correct' answer in that the actual outcome of the scenario was a try (the optimal outcome). The scene was cropped to remove identifying information to ensure that players did not recognise the particular scene.

Figure 1. Critical Decision Method Picture Stimulus

Data analysis

Data was analysed using a structured methodological approach in the form of a content analysis. Content analysis is a technique which provides knowledge and understanding of a phenomenon which ensures that all units of analysis receive equal treatment (Krippendorff, 2012). Participants were categorised and aggregated based on their level of performance and their experience. A coding framework was developed for the purpose of identifying cue use in order to guide the directed content analysis.

Text was also segmented into clusters and concepts which informed the decision maker to create concept maps. This was based on the methodology of Glaser, Lesgold, and Lajoie (1987). Concept Maps were used to represent the knowledge structures of the different categories of players as well as contextualise the cues and their constituted relationships. Concept maps have been used widely in the study of expertise (Cañas et al., 2005), often showing significant differences between experts and novices in knowledge and the structure of that knowledge (Glaser et al., 1987).

A cognitive demands table was used to decompose tasks in order to identify cue based behaviour and allow for comparison across category groups (Militello & Hutton, 1998). The text was analysed through the identification of difficult cognitive elements and the problems and methods used by players to overcome them as outlined by Militello and Hutton (1998).

Critical cue inventories were used to organise the informational and perceptual cues that were present during a given protocol (Klein, 1996). Content analysis techniques were used to construct a critical cue inventory. This involved identification of cues based on previous operational definitions and making determinations as to whether they were present in each of the different categories of participants.

Cues were identified and tallied. If the player continued to refer to the cue that initiated his response it was only counted as one cue, consistent with Baber and Butler (2012). Parametric analysis was considered inappropriate due to the small sample size.

RESULTS

Cue Count

A cue count was conducted from both the cognitive task analysis and the picture stimulus (Table 2).

Table 2. Mean Cue across Category of Players

Category of Player	Mean Cue Count
4	39
3	28
2	35
1	24

Concept map

A concept map was created to analyse the relationship between cues used during decision-making and the decision response, and to compare across the different categories of players. It was observed that experts place different weighing and meaning on different cues. Concept maps allowed for this different representation of knowledge to be analysed. Cue correlations and structural differences were observed in the representation of knowledge as a function of category of player. See Figure 2 for concept map.

Cognitive Demands Table

There existed cognitive elements for which could be measured to show levels of knowledge and experience. The results of this breakdown of elements are presented in Table 3.

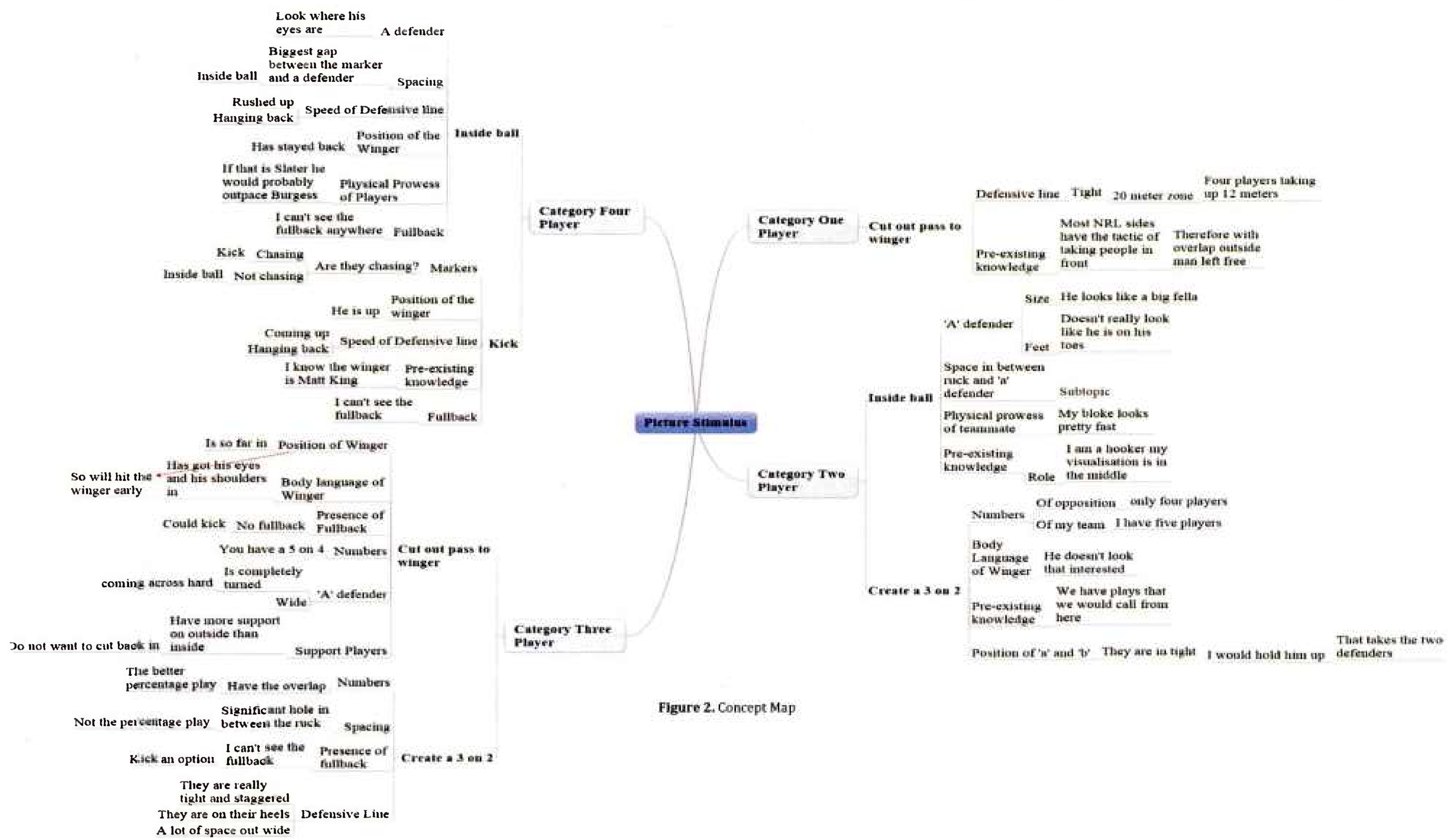


Figure 2. Concept Map

Table 3. Cognitive Demands Table

Difficult Cognitive Element	Why Difficult?	Common Errors	Cues and Strategies Used
Picture Stimulus	There are a lot of options and a lot of unknowns	Going away from your support Not playing to your players capabilities	Player cues: Where he is looking, size, shoulders, eyes Numbers "When you have the numbers you let the defenders make the decisions" How compact the defence is "Within the space of this 20 metre zone here they are taking up 12 of it" Timing "Not showing your hand too early" Space "Biggest gap is in between the marker and the a defender" Ability of your players "He knows that **** is the centre and knows that if **** gets the ball he will score"

Critical Cue Inventory

In the *Picture Stimulus*, a total of 10 critical cues were identified (Table 4). Differences in the presence of cues as a function of category of player were observed. For Category 4 players, $N = 5$, Category 3 players $N = 7$, Category 2 players $N = 7$, Category 1 players $N = 2$.

Table 4. Critical Cue Inventory

Cognitive Element	Cue	Descriptor	Category of player 1	2	3	4
Picture Stimulus						
	Space	"There is a big gap behind the ruck"	✓	✓	✓	×
	Numbers	"You obviously have the overlap, you would look at the numbering which is in our favour"	×	✓	✓	×
	Position of players	"Winger is just banging back sort of thing and he has got bis eyes and his shoulders in sort of thing so probably looking to hit the centre or the winger early" "I can't see the fullback so the kick would be an option"	✓	✓	✓	×
	Size	"That looks like a big fella so I would go to the outside of A"	✓	×	✓	×
	Eyes	"I would look where his eyes are, if he is looking at my support player"	✓	✓	✓	×
	Shoulders	"I am looking at their shoulders whether the shoulders are turned in or out"	×	✓	×	×
	Defenders	Whether they have rushed up or are hanging back"	✓	×	✓	×
	Support players	"You can't really take the line on here well you can but because you have 5 on 4 you don't really want to come back to a because that is just away from your support"	×	✓	×	×
	Defensive Line	"Within the space of this 20 metre zone here they are taking up 12" "They are really tight and you can see the staggered line"	×	✓	×	✓
	Feel	"I could see centre was either flat footed or his feet stayed planted"	×	×	✓	✓
			5/10	7/10	7/10	2/10

DISCUSSION

As anticipated, cue use was found to be an important aspect of playing the game of Rugby League. All players recognised that cues are an effective source of information. Whilst some cues were used universally across all categories of players, there were important differences with respect to the ability for lower category players to discriminate amongst cues. There existed differences in the meaning attributed to the cues. There also existed differences in how cues were used. Higher category players were able to recognise and use cues whilst low category players were able to discriminate between cues as well as having the capacity to control the opposition through deliberate moves to misdirect their cue reading (counter cues).

Cue Use and Rugby League

Consistent with previous research, cues were used as a means to inform decision making. Players used body cues such as position of the shoulders or where the player was looking as a means to predict player movement (see Table 3). This is consistent with Macquet and Fleurance (2007) and Renshaw and Fairweather (2000) who showed that players control the situation by focusing on salient cues which allow them to make more accurate predictions.

Whilst individual cues were identified, analysis revealed that the association between separate cues as components of a correlated relation was a factor. In the performance of a kick, each category of player recognised the importance that the position of the winger in conjunction with the position of the fullback held.

A Category two player identified this, "The main thing is just the position of the full back and the winger". This suggested that players do not just use cues in isolation, but rather, it is the association between each of the cues which activates the conditioned-action response. Morrison, et al. (2013) suggested that cues which are correlated together form cognitive links. These links result in a reduction in amount of cognitive resources used rather than if the individual had to consider the cues separately. Wickens and Hollands (2000) hypothesised that these individuals were able to create higher order cognitive representations of items within long term memory structures. The theoretical implication of this is that it leads to building structures upon structures, the problem of infinite regress.

Consistent with previous research, players associated with a higher degree of expertise showed greater cue discrimination (Shanteau & Hall, 1992). A comparison between a Category One and a Category Three player on the picture stimulus task who both gave the 'correct' answer showed differences in their cue use. A cue count revealed that the Category three player required 6 cues compared to the 2 cues required by the Category one player. This isolated effect is reflective of the pattern observed globally, which indicated that the number of cues used decreased as a function of level of expertise. This is consistent with findings which have shown that highly skilled operators only engage a limited number of critical cues, as opposed to less skilled operators whom reportedly engage a number of non-relevant cues (Martell & Vickers, 2004; Raab & Johnson, 2007). It may be that experts are able to 'chunk' more information than novices with no loss in the amount of information conveyed in those 'chunks' consistent with observations made by Chase and Simon (1973). This could also be explained through pattern recognition skills, which have been shown to be more prevalent in experts (Abernethy, Baker, & Côté, 2005).

Expertise within a Rugby League context is more than just the performance of a skill it also involves a refinement in the information processing of cues. The suggestion that the environment plays an important role in refining the cue discrimination process does not fully account for the ability of some players to be an active agent in this process. Some insight into this may be addressed through the way that players attribute meaning to the cues.

Meaning of the Cues

Previous research has shown that experts ascribe different meaning to cues which produced quantitatively different outcomes (Crandall & Getchell-Reiter, 1993). Consistent with this research, the players interviewed showed differences in the meaning they assigned to cues. Category four and Category one players both recognised size as an important cue. Category four players assigned the meaning that size displayed a negative relationship with speed of the player such that a big player is thought of as slow. This is contrasted with the Category one player who assigned the additional meaning that size of the player showed a negative relationship with the speed of the play-the-ball. This discrimination within the cue creates differences in the retrieved representation and the associated response in its activation. Such that for the Category one player, this meant actively avoiding big players in contrast to Category four players who took that cue as a signal to attack. The recognition of this difference in meaning attributed by the category four and category one players suggested that creating a critical cue inventory for assisting in training is not sufficient. Any training program which would be created would need to make explicit this difference in meaning.

The difference in the category one's reporting of how he used his previous experience to inform his cue use was markedly different to how other categories assigned meaning and importance to the cues used. Category one player "...most NRL sides are likely to have the tactic of taking the people in front of you and leaving the man free so to speak.." therefore "I would identify how tight they are so within the space of this twenty meter zone here". A Category 3 player "you don't really want to come back to 'A' because that is just away from your support," therefore you would identify options outside of you. This is in contrast to a category 4 player "the biggest gap is in between the marker and the 'a' defender so I would probably go for the inside ball" who has assigned the cue with the greatest meaning to be the gap instead of his support players. Players responded to the cues they had identified as most important. It has been suggested that experts are able ascribe meaning and direct salience to the most effective cues (Klein, Calderwood, & Clinton-Cirocco, 1986).Whether the capacity to do this is the product or a result of expertise is yet to be determined. It was interesting that the category one player and category three player both came to the same decision yet relied on a different set of cues.

How Cues Were Used

Cues used in decision making may not always trigger automatic associations but can trigger rule based responses (Greitzer et al., 2010). A Category four player responded to the information available within a structured rule based way, Category four player "I just pictured their being two defenders a lot wider than usual... and just pictured that and mentally rehearsed that so I could jump out and easily do that", this is in contrast to a category one player who reacted to the information in an automatic, subconscious level without the need to interpret and integrate the information, "before catching the ball my mindset was a lot different to when I caught the ball... from my vision going from the defensive line to me catching the ball and then looking back and noticing that they had adjusted differently then changed my thoughts from not just running the ball but possibly getting around people." This is consistent with Greitzer et al. (2010) who observed that novices process input and perform tasks at a rule based level whereas experts were able to respond at a skill based level, reacting automatically with little conscious thought. This reflected the observed progression in cue use with lower category players showing an ability to recognise that they were an active agent who exhibited cues and therefore would influence the cues that they would project to deceive other players. "I am enticing players to come out of the line. I've got the ball out in two hands and dummying, so with that being done they are obviously, they are not thinking that you are going to kick it." This was an ability not observed in higher category players who only identified that cues could be used as an important source of information. It is suggested that in freeing up cognitive resources more complex cue clustering may occur. This could be one explanation for the observation that experts have more complicated and detailed knowledge structures (Gourlay, 2006).

The change in cue use seems to suggest a complex process. A more in-depth exploration of the triggers, the variations, the decision thresholds, weightings and cue clusters are required. It would appear that players have developed some of these meanings and weightings based on previous experience. It could be hypothesised that the higher category players are continuing to reorder the cues significance whereas the lower category player is confident in the reliability and validity of his cue choice to provide the outcome he wants. This is an area which requires future research.

Applications and Future Research

The methodological focus on cues meant that other aspects which may have further differentiated expert performers were not taken into account. These would include playing style or creativity, sometimes labelled instinctual playing. It would be helpful to develop questions around the players' use of cue discrimination and the reliability of cues used. This may provide greater understanding of how this cue use is developed.

An insight provided by a Category two player suggested that the environment plays an important role in cue discrimination. Cue-based training programs work on this assumption (Perry, Wiggins, Childs, & Fogarty, 2013) and may be an effective means for novices to increase cue discrimination and learn reliable cue association. The results of this study suggested that all players have knowledge of the array of available cues, but differences exist in the number of cues they rely on. Zsambok (1997) developed a cue-based training program that encouraged the user to focus on and use critical cues associated with a task. This may be a more appropriate way of developing training although these training programs so far have not been applied to sporting domains. However, an important consideration is ensuring that the meaning of the cue and not simply the identification of critical cues is trained. Using virtual reality technology to develop cue association simulations may provide possibilities for future training.

References

Abernethy, B., Baker, J., & Côté, J. (2005). Transfer of pattern recall skills may contribute to the development of sport expertise. *Applied Cognitive Psychology, 19*(6), 705-718.

Baber, C., & Butler, M. (2012). Expertise in crime scene examination comparing search strategies of expert and novice crime scene examiners in simulated crime scenes. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 54*(3), 413-424.

Barton, J. (2013). Soward asked to find next Penrith NRL star, *The Sydney Morning Herald.*

Cañas, A. J., Carff, R., Hill, G., Carvalho, M., Arguedas, M., Eskridge, T. C., . . . Carvajal, R. (2005). Concept maps: Integrating knowledge and information visualization *Knowledge and information visualization* (pp. 205-219): Springer.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive psychology, 4*(1), 55-81.

Crandall, B., & Getchell-Reiter, K. (1993). Critical decision method: A technique for eliciting concrete assessment indicators from the intuition of NICU nurses. *Advances in Nursing Science, 16*(1), 42-51.

Gabbett, & Abernethy, B. (2013). Expert–Novice Differences in the Anticipatory Skill of Rugby League Players. *Sport, Exercise, and Performance Psychology, 2*(2), 138-155.

Glaser, R., Lesgold, A., & Lajoie, S. (1987). Toward a cognitive theory for the measurement of achievement.

Gourlay, S. (2006). Towards conceptual clarity for 'tacit knowledge': a review of empirical studies. *Knowledge Management Research & Practice, 4*(1), 60-69.

Greitzer, F. L., Podmore, R., Robinson, M., & Ey, P. (2010). Naturalistic decision making for power system operators. *Intl. Journal of Human–Computer Interaction, 26*(2-3), 278-291.

Klein. (1996). The Development of Knowledge Elicitation Methods for Capturing Military Expertise: DTIC Document.

Klein, Calderwood, R., & Clinton-Cirocco, A. (1986). *Rapid decision making on the fire ground.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society annual meeting.

Krippendorff, K. (2012). *Content analysis: An introduction to its methodology*: Sage.

Macquet, & Fleurance, P. (2007). Naturalistic decision-making in expert badminton players. *Ergonomics, 50*(9), 1433-1450.

Martell, S., & Vickers, J. (2004). Gaze characteristics of elite and near-elite athletes in ice hockey defensive tactics. *Human Movement Science, 22*(6), 689-712.

Militello, L., & Hutton, R. (1998). Applied Cognitive Task Analysis (ACTA): A practitioner's toolkit for understanding cognitive task demands. *Ergonomics, 41*(11), 1618-1641.

Morrison, Wiggins, Bond, & Tyler. (2013). Measuring relative cue strength as a means of validating an inventory of expert offender profiling cues. *Journal of Cognitive Engineering and Decision Making*, 1555343412459192.

Perry, Wiggins, Childs, & Fogarty. (2013). The Application of Reduced-Processing Decision Support Systems to Facilitate the Acquisition of Decision-Making Skills. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 55*(3), 535-544. doi: 10.1177/0018720812467367

Proszenko, A. (2013). Academy just part of Penrith's bold plan to return to their glory days, *The Sydney Morning Herald*.

Raab, M., & Johnson, J. G. (2007). Expertise-based differences in search and option-generation strategies. *Journal of Experimental Psychology: Applied, 13*(3), 158.

Ratcliff, R., & McKoon, G. (1994). Retrieving information from memory: spreading-activation theories versus compound-cue theories.

Renshaw, I., & Fairweather, M. M. (2000). Cricket bowling deliveries and the discrimination ability of professional and amateur batters. *Journal of sports sciences, 18*(12), 951-957.

Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes, 53*(2), 252-266.

Shanteau, J., & Hall, B. (1992). How Much Information Does An Expert Use? Is It Relevant?

Shim, J., Carlton, L. G., Chow, J. W., & Chae, W.-S. (2005). The use of anticipatory visual cues by highly skilled tennis players. *Journal of motor behavior, 37*(2), 164-175.

Tenenbaum, G., & Bar-Eli, M. (1993). Decision making in sport: A cognitive perspective. *Handbook of research on sport psychology. New York*, 171-192.

Wiggins. (2006). Cue-based processing and human performance. *Encyclopedia of ergonomics and human factors*, 641-645.

Yarrow, K., Brown, P., & Krakauer, J. W. (2009). Inside the brain of an elite athlete: the neural processes that support high achievement in sports. *Nature Reviews Neuroscience, 10*(8), 585-596.

Zsambok, C. E. (1997). Naturalistic decision making research and improving team decision making. *Naturalistic decision making*, 111-120.

Clear and Present Danger: A Virtual-Reality Cue-Based Training System to Aid Police in the Detection of Imminent Terrorist Attacks

Ben MORRISON[ab], and Natalie MORRISON[ac]
[a]*Australian College of Applied Psychology*
[b]*Macquarie University*
[c]*University of Western Sydney*

ABSTRACT

Since the 2001 terror attacks in the United States, counter-terrorism strategies have become a security priority for many nations. More recently, the upsurge of 'home-grown' extremists and those returning home from foreign conflicts represents a clear and present danger to public safety. Unfortunately, the pre-identification of would-be terrorists is an extremely difficult task for law enforcement, especially in the final stages of a planned attack. The authors propose that training programs designed to improve law enforcement threat detection may benefit from simulations which incorporate expert knowledge relating to terror indicators. The current paper details research in progress involving the development of a virtual-reality cue-based training program, which simulates environments that are presumably vulnerable to terror attacks (e.g., public transport), and cues police trainees to environmental features and behaviours that may indicate an impending attack.

KEYWORDS

Learning and Training; Security; Decision Making; Expertise; Government and Law

INTRODUCTION

Since the 2001 terror attacks in the United States, counter-terrorism strategies have become a security priority in many nations. More recently, the upsurge of 'home-grown' terrorists and those returning home from foreign conflicts represents a clear and present danger to public safety. Indeed, in a number of countries, the government-directed threat assessment level is currently at 'high'; meaning that an attack is likely (National Terrorism Public Alert System, Australian Government; 2015).

A terrorist attack is often likened to a missile launch, in that once the missile has been launched and is approaching a target, it is extremely difficult to thwart. Indeed, without intelligence, identification of a would-be terrorist during the final stages is an ominous prospect for law enforcement. This challenge is of course compounded by the limited avenues for stopping the attack, once an offender has been identified. Thus, much research has focussed less on situational prevention and more on the 'launch' phase. That is, detection of terrorist activities at the prepatory stages of an attack (e.g., reconassence, rousourcing, 'dry runs', etc.), using approaches such as *Social Network Analysis* modelling (Ressler, 2006), in an effort to improve intelligence-based procedures.

Despite this emphasis on the intelligence process, Police undoubtedly remain the front line defence to terrorism. Indeed, it is estimated that 90 per cent of information on potential terrorist threats comes from local law enforcement (U.S. Department of Homeland Security, 2003). The capacity for police to assess scenes rapidly and accurately, particularly those involving large volumes of people (e.g., public transport, major events) is paramount to preventing an attack, or minimizing its impact. Currently, many police organisations can only provide officers limited specialised training in threat assessment and the initiation of countermeasures should detection occur.

When considering training design for such a task, the Naturalistic Decision Making (NDM) paradigm may offer a useful approach. Traditionally, NDM research has sought to improve human performance by modeling the behaviours and cognitive processes employed by experts when formulating successful decisions. On the basis of this research, there is an increasing body of evidence to suggest that observed differences in situational assessment skills are differntiated by the capacity to target and engage relatively diagnostic *cues* present in the environment (Wiggins, 2006; Klein, 2004). From a cognitive perspective, such cues are presumed to represent relationships, held in memory, between environmental or situational features, and consequential events of interest (Wiggins, Azar, Hawken, Loveday, & Newman, 2014).

The use of cues has been shown to differ across expertise, to the point where highly experienced decision-makers may only require a limited number of cues to formulate a decision (Wiggins, 2014). Indeed, in numerous domains such as fire-fighting (Klein, Calderwood, & MacGregor, 1986), medical diagnoses (McCormack, Wiggins, Loveday, & Festa, 2014), and offender profiling (Morrison, Wiggins, Bond, & Tyler, 2013), highly proficient operators have been shown to consistently engage a limited number of *critical* cues across varying decision scenarios. In comparison, less experienced practitioners tend to engage cues less consistently across decision scenarios (Boreham, 1995; Kirschenbaum, 1992; Morrison et al., 2013). Arguably, it is this distinction between expert' and novice' application of cues that contributes to the less rapid and less accurate processes observed in novice decision-making. As such, the use of a sample of cues commonly engaged by proficient/expert operators (i.e., an expert critical cue inventory) is an attractive avenue for skill development. These cues may be extracted using Cognitive Task Analysis and eye-tracking techniques, and may be validated and refined to a critical set using cue recognition tasks (Morrison et al., 2014).

The authors postulate that such cue-based training methods may augment current police training in the rapid assessment of scenes for threats to security. Indeed, recent evidence has heralded the benefits of *guided discovery* in skill acquisition in anticipatory tasks (Smeeton, Williams, Hodges, & Ward, 2005). Guided discovery involves constraining the learner's focus of attention to relevant cues, and is said to be an improvement upon more prescriptive approaches which invariably increase load, slow acquisition, and result in less robust learning than more implicit approaches (Green & Flowers, 1991).

Expert knowledge regarding such threat detection has been successfully extracted and has been used by law enforcement largely in a 'ckecklist' manner. For instance, Israeli Police have performed a significant number of interviews with witnesses of suicide bombers in order to identify the features that may indicate an impending attack (see Table 1).

Table 1.A list of indicators published by Israel Police in 'Terror: Let's Stop it Together', which was based on knowledge drawn from interviews with suicide bomb experts.

External appearance
- Clothes unsuitable for the time of year (e.g., a coat in summer).
- A youngster (usually) who is trying to blend, by dress and behavior, with the surrounding population (on public transport, at entertainment places, amongst soldiers, or religious/Orthodox groups), even though he or she doesn't belong to that group.
- Anything protruding in an unusual way under the person's clothing.
Suspicious behavior
- Nervousness, tension, profuse perspiration.
- Walking slowly while glancing right and left, or running in a suspicious manner.
- Repeated attempts to steer clear of security forces.
- Repeated nervous feeling for something under one's clothing.
- Nervous, hesitant mumbling.
Suspect equipment
- A suitcase, shoulder/hand-bag, backpack.
- Electrical wires, switches or electronic devices sticking out of the bag or pocket.

These indicators, or what NDM researchers may term critical cues, could be embedded in high fidelity simultaion programs now readily available to training system developers (e.g., Oculus Rift). For instance, the critical cues mentioned here could be embedded into a training simulation featuring a busy commuter train, or a crowded shopping centre. Guided discovery cueing techniques could then highlight features of interest for trainees, as they navigate these immersive environments.

The proposed simulations may offer a convenient (e.g., infinitely replayable scenarios; exposure to a range of conditions, including non-routine cases), cost-effective, and safe approach to simultating high stakes environments that would be difficult to stage in the physical world. Investigating the efficacy of these techniques presents a promising avenue for augmenting current approaches to training for threat detection, which are presently quite limited.

RESEARCH QUESTION

The authors propose that training programs designed to improve threat detection may benefit from the embedding of expert knowledge and guided discovery techniques. The authors are developing a virtual-reality cue-based training program, which simulates environments that are vulnerable to terror attacks (e.g., public transport), and cues police trainees to environmental features that may indicate an impending terrorist attack. Several studies are planned to evaluate the efficacy of this training program against other traditional methods (e.g., indicator checklists) and pure discovery learning.

METHOD

Materials and Procedure

3D Modelling, Rigging, and Animating software (UnityPro 4) will be used to create several virtual environments to be run on a Oculus Rift virtual reality headset (Figure 1) and Virtuix Omni Walker. Participants will be fully immersed in the experience of a patrolling a public scene for threats to public safety. These simulations will be augmented with animations to highlight key regions of the visual scene, which may be relevent to threat identification. See Figure 2 of a projected composite of the training experience/conditions.

Figure 1

Figure 2

REFERENCES

Boreham, N. (1995). Error analysis and expert-novice differences in medical diagnosis. In J-M Hoc, P.C. Cacciabue, & E. Hollnagel (Eds.), Expertise and technology (pp. 93-105). Hillsdale, NJ: Lawrence Erlbaum.

Green, T. D., & Flowers, J. H. (1991). Implicit versus explicit learning processes in a probabilistic, continuous fine-motor catching task. Journal of Motor Behavior, 23, 293–300.

Kirschenbaum, S.S. (1992). Influence of experience on information-search strategies. Journal of Applied Psychology, 77, 343-352.

Klein, G. (2004). The power of intuition. New York: A Currency Book/Doubleday.

Morrison, B. W, & Wiggins, M. W., Bond, N., & Tyler, M. D. (2013). Relative Cue Strength as a Means of Validating an Inventory of Expert Offender Profiling Cues. Journal of Cognitive Engineering and Decision Making, 7(2), 211-226.

Ressler S. (2006). Social network analysis as an approach to combat terrorism: Past, present, and future research. Homeland Security Affairs,2, 1-10.

Smeeton, N. J., Williams, A. M., Hodges, N. J., Ward, P. (2005). The relative effectiveness of various instructional approaches in developing anticipation skill. Journal of Experimental Psychology: Applied, 11(2), 98-110.

U.S. Department of Homeland Security (2003), Potential Indicators of Threats Involving Vehicle Borne Improvised Explosive Devices. Homeland Security Information Bulletin

Wiggins, M.W. (2014). Differences in situation assessments and prospective diagnoses of simulated weather radar returns amongst experienced pilots. International Journal of Industrial Ergonomics, 44(1), 18-23. http://dx.doi.org/ 10.1016/j.ergon.2013.08.006

Wiggins, M. (2006). Cue-Based processing and human performance. In I. W. Karwowski (Ed.), Encyclopedia of ergonomics and human factors (pp. pp. 3262 - 3267). London, UK: Taylor and Francis.

Wiggins, M.W., Azar, D., Hawken, J., Loveday, T., & Newman, D. (2014). Cue-utilisation typologies and pilots' pre-flight and in-flight weather decision-making. Safety Science, 65, 118-124.

Identifying Critical Cues in Mental Health Assessment using Naturalistic Decision-Making Techniques

Ben MORRISON[ab], Julia MORTON[a], and Natalie MORRISON[ac]
[a]*Australian College of Applied Psychology*
[b]*Macquarie University*
[c]*University of Western Sydney*

ABSTRACT
Decision cues are important components of situation assessment. The identification of seemingly critical cues has proven beneficial to training initiatives in a number of domains (e.g., fire-fighting, aviation, nursing, criminal investigation). Similarly, it is proposed that a critical cue inventory may augment training opportunities in the mental health domain (e.g., developing high-fidelity virtual patients). To date, there has been no formal identification of the cues engaged by Mental Health Practitioners (MHP). This study used the Critical Decision Method to decompose the initial stages of psychological assessment, and elucidate the cues engaged by practicing MHPs. Further, it examined MHPs' perceptions of diagnosticity (i.e., predictive value) and frequency of use, to identify those cues most critical to assessment. The results reveal that MHPs engage an array of cues, and an inventory of critical cues is presented. Findings may be used to inform training and decision support initiatives in clinical skill acquisition.

KEYWORDS
Decision Making; Health; Situation Assessment; Cognitive Task Analysis

INTRODUCTION

For Mental Health Practitioners (MHPs), decision-making is integral to competency (Vollmer, Spada, Caspar, & Burri, 2013). Decision-making in this domain requires that the MHP encode, manipulate and recall information to formulate decisions for tasks including assessment, case conceptualisation, diagnosis, risk assessment, and treatment planning for clients (Whaley & Geller, 2007). These decisions are frequently high stake, time-pressured, and uncertain, due to the complexity inherent in psychological problems.

For the MHP, decision processes often differ from physiological medical assessments in that there are rarely overt symptom data available for consideration (Broadbent, Moxham, & Dwyer, 2007), and the data on which decisions are based are often self-report with uncertain validity (Bhugra, Easter, Mallaris & Gupta, 2011). Additionally, attempts to investigate and understand the cognitive processes involved in MHP decision-making are complicated by frequent claims that clinical decisions are based largely on intuition, particularly in mental health nursing (King, 1997), general mental health care (Wittemann, Spaanjaars, & Aarts, 2012), and psychiatry (Bhugra et al., 2011).

The role of intuition in decision-making is regularly discounted in the literature, most likely due to a perceived lack of empirical evidence (Grove, Zald, Lebow, Snitz, & Nelson, 2000). However, there is evidence to suggest that MHPs combine both intuition and empirical methods in clinical decision-making, recognising that intuition produces hypotheses that require formal validation. Indeed, Welsh and Lyons (2001) offer that intuition is the combination of the MHPs formal knowledge, coupled with experience to create a store of tacit knowledge from which to draw upon during decision-making. This may offer explanation as to why MHPs are able rationalise their decisions in diagnostic assessment in the absence of formal assessment psychometrics. The increasing recognition of intuition as a valid process within complex decision-making has seen a shift in decision-making research from the prescription of systematic optimization-based strategies, to the examination of decision-making in real-world or naturalistic settings.

Naturalistic Decision-Making (NDM) Paradigm

The NDM framework has shifted the extant conception of decision-making from one of a domain-independent and generalised approach, to that of a knowledge-based approach incorporating the individual decision-maker's previous experiences and stores of knowledge (Klein, 2008). Klein and Klinger (1991) suggest that decision-making processes may be represented on a continuum from the analytical strategies at one end to the recognition-based decision strategies at the other, and strategy engagement fluctuates between these extremities, depending on the nature of the situation. Within the NDM paradigm, processes of decision-making are extended to include constructs such as the prior stages of perception and situational recognition, in addition to the notion that individuals generate relevant responses rather than simply choosing from a given set of responses (Klein, 2008).

NDM research tends to emphasise the cognitive processes that contribute to a decision-maker identifying effective courses of action. For instance, in a range of domains, practitioners' ability to trigger meaningful associations in memory by identifying relevant environmental indicators (i.e., cues), appears to be a key differentiation in decision-making performance (Beilock, Wierenga, & Carr, 2002; Klein, 1993; Loveday, Wiggins, & Searle, 2013; Morrison, et al., 2013; Schriver et al., 2008 ; Perry, Wiggins, Childs, & Fogarty, 2013). As a result, cue use is viewed as a prominent avenue of interest for NDM researchers looking to model proficient processes in training programmes.

Cues and Cue Diagnosticity

Cues have been found to be crucial in decision-making performance across a number of domains including medical diagnoses (Hammond, Frederick, Robillard, & Victor, 1989), courtroom judgments (Ebbesen & Konecni, 1975), aviation (Stokes, Kemper, & Marsh, 1992), airport customs (Pachur & Marinello, 2013), power control (Loveday et al., 2012), finance (Hershey, Walsh, Read, & Chulef, 1990), driving (Fisher & Pollatsek, 2007), nursing (Shanteau, 1991), and criminal investigation (Morrison et al., 2013). Much of the research examining cue use has been in domains with time constraints, high information load and serious consequences, such as power control and aviation (Loveday et al., 2012; Wiggins & O'Hare, 2003), where effective performance implies the rapid assessment of the situation to reach accurate decisions within a specified time frame. Although decisions in mental health are not necessarily rapid, Schmidt and Boshuizen (1993) suggest that in the health care domain, cues still play an important role, and are probably reflected in associations between diagnostic features and patient events or symptoms that are stored in the long-term memory of the practitioner.

As evidence for the importance of cues in accurate decision-making is mounting, more research is focussing on designing training initiatives that promote cue discovery. For example, Wiggins and O'Hare (2003) developed a computer-based training system designed to enable pilots to identify critical cues associated with deteriorating weather conditions during flight. The aim of this type of training has been to expose the learner to cues that are useful as triggers for diagnosis. One promising application for cue-based training in the mental health domain is the development of virtual patient technologies used in psychological assessment. For example, Kenny, Parsons, Gratch, and Rizzo (2008) have developed a virtual patient, Justina, designed to portray a victim of sexual assault, communicating symptoms of Post-Traumatic Stress Disorder during a clinical interview. From this simulation of the assessment process, trainee MHPs learn to formulate preliminary hypotheses and diagnoses. It is proposed that one way to improve these simulations would be to enhance their capacity to demonstrate more subtle indicators of symptomology that are invariably engaged by MHPs during practitioner-patient interactions.

Aim

The aim of this study was to identify a critical cue inventory utilised by experienced MHPs from a range of practicing approaches for use in future training initiatives. This study sought to achieve this by (a) using the critical decision method with a number of experienced MHPs to extract a range of cues used during the initial stages of psychological assessment; and b) using a survey, investigate whether there are significant differences in MHPs ratings of perceived diagnosticity and frequency of use across the cues extracted, to determine the most critical cues.

METHOD

Participants

The participants comprised two separate purposive samples. Firstly, 12 mental health professionals; five were practicing registered psychologists, four practicing clinical psychologists, one practicing registered counsellor, one practicing forensic psychologist and one practicing registered social worker. Participants ranged in age from 31 to 57 years (M_{age} = 42.16, SD =9.25) and years' experience ranging between 6 to 15 years (M = 10.14, SD =4.38). Based on these factors, it was believed that this sample would produce a rich and diverse range of knowledge and skills, likely not possess by training MHPs. Secondly, 50 mental health professionals participated in and completed the online survey advertised in the Australian Psychological Society Bulletin. There were 40 female participants and 10 male participants. The mean number of years practicing as a MHP was 8.8 years.

Materials and Procedure

The initial 12 participants participated in a 60-minute audio recorded, semi-structured interview. The interview schedule was based on a form of Cognitive Task Analysis - the Critical Decision Method (CDM) procedure adapted from Klein, Calderwood and Macgregor (1989) - which can be used to elicit cues during decision-making (see O'Hare, Williams, Wiggins, & Wong, 2000 for further detail).

Participants were asked to recall (retrospectively) and recount to the interviewer details pertaining to a non-routine case that they had assessed. Here it was assumed that the use of non-routine cases were more likely to involve greater intricacy, offering a richer source of data for analysis and elicit tacit knowledge stores of the domain expert (Crandall, et al., 2006). Interviews were transcribed for protocol analyses.

The next 50 participants were invited to complete an online survey designed to assess their perceptions of diagnosticity (i.e., predictive value) and frequency of use of the cues extracted from the interviews. Participants were presented with each of the 73 cues and asked to make two ratings for each regarding the two dimensions of interest; diagnosticity (i.e., How *relevant/important* are the following cues to assist in the assessment of your client's mental health status?) and frequency of use (i.e., How *frequently* do you rely on/utilise the following cues to assist in the assessment of your client's mental health status?).

RESULTS

Critical Decision Method

The transcribed data was analysed for the abstraction of cue-based information and decomposition into content-based coding categories. Critical Task Analysis (CTA) offers numerous coding schemes that are established based on the task domain and the purpose or goal of the analysis (Crandall et al., 2006).

The categories selected to represent the task of initial assessment were based on descriptions offered by Crandall et al. (2006) and include informational cues, hypothesis formation, hypothesis testing, seeking information, sense making, mental models, and reference to knowledge.

The first level of protocol analysis was based on a content abstraction process to identify information that was relevant to the population of these categories. The aim here was to assign each interviewee's relevant verbalizations to one of the following categories; informational cues, hypothesis formation, hypothesis testing, seeking information, sense-making, mental models, and reference to knowledge. This process yielded a number of what could be described as content (e.g., medication) and perceptual (e.g., tone) cues.
The second level of protocol analysis narrowed the focus to cues further, and involved higher level coding; collapsing the cues identified in the first level of abstraction into thematic categories. For example the perceptual cues of pitch, tone, pauses, and volume were all collapsed into the general content category of *Speech*, while cues of hand gestures, fidgeting and threatening stance were summarized by the category *Body Movement*.

Overall, 73 individual cues, and 11 cue-categories were extracted from the protocol analysis. Table 1 shows a selection of cues and their respective categories.

Table 1. The 11 Cue Categories and examples of specific cues from each

Cue Category	Examples of Cues Included
Personal Information Cues	Gender, occupation, race, religious affiliations, socioeconomic status, appearance, lifestyle factors
Medical Cues	Medication prescribed, compliance, blood serology, previous diagnosis, current diagnosis, family history of diagnoses
Immediacy Cues	Engagement, affect, communication style, facial expressions, emotional expression, personality traits/temperament, transference
Speech Cues	Tone, flow, perseverative, slurred, volume, pitch, pace
Language Cues	Descriptors, words used, developmentally appropriate, use of humour
Physical Cues	Breathing, eye contact, voice, body movements
Cognitive Cues	Attention, memory, intelligence, intellectual disability, judgement, decision making,

	perceptions
Risk Assessment Cues	Risk of harm to self, others, intent, means and plan
Collateral information Cues	Congruency between verbal and non-verbal, consistency between collateral, psychometrics and narrative Referral source and question
Overt Behavioral Cues	Behaviour in waiting room, occupation of space in therapeutic environment, feedback from client
Personal History Cues	Psychosocial history, relationship status, conflicts, support networks

Cue Survey

Analysis of the survey data involved two phases. Firstly, to investigate whether there were significant differences in participants' perceived diagnosticity and frequency of use for the cues, and secondly, to identify those cues with the highest ratings of diagnosticity and frequency (i.e., the most critical cues).
As the large volume of cues represented a challenge to statistical comparison, ratings from each cue were collapsed into their respective categories, resulting in grand ranked means. The assumption of normality for parametric analysis was not met for several cue categories.

Two Freidman's tests were used to determine whether significant differences existed in; 1) the frequency with which participants rated their perceived use of the cues within each category; and, 2) the perceived diagnosticity (i.e., operational relevance) of each cue category.

Firstly, a Freidman's test compared ranked means for frequency of use for each of the 11 cue-categories. With alpha set at .05, the results revealed a statistically significant effect, χ^2 (10) = 210.93, $p < .001$. Post hoc comparisons were performed between pair-wise means using Wilcoxon Signed-Rank tests, and a Bonferroni adjusted alpha of .001. Significant differences in perceived frequency were found between 36 of the 55 comparisons. Notably, the means for frequency of cues from the *Risk* category were significantly higher than means from all other cue categories.

Secondly, a Freidman's test compared ranked means for diagnosticity for each of the 11 cue categories. Alpha was set at .05. The Friedman's test was significant, χ^2 (10) = 220.74, $p < .001$. Post hoc comparisons were performed between pair-wise means using Wilcoxon Signed-Rank tests, and a Bonferroni adjusted alpha of .001. Significant differences in perceived diagnosticity were found between 31 of 55 comparisons. Of note, means for perceptions of diagnosticity for *Risk* cues were significantly higher than means from all other cue categories, which is consistent with participants' perceptions of frequency for these cues.

To affirm the assumed relationship existed between participants' perceptions of frequency and diagnosticity, Spearman's Rho correlations were conducted between means for each cue category's frequency and the means for each cue category's diagnosticity. With alpha set at .05, strong (r = >.5), positive, and significant correlations were found between frequency and diagnosticity for each cue category.

Finally, to identify those cues with the highest ratings of diagnosticity and frequency (i.e., the most critical cues), mean ratings of diagnosticity and frequency for each cue were combined, and cues with a combined mean rating of greater than four (i.e., Very relevant/Almost always used) were retained as the most critical cues from the sample. The critical cue inventory is shown in Table 2. As a result of this process, 28 of 73 (38%) of the cues extracted were deemed to be critical to the practitioner sample. Further, consistent with the previous analyses, the most critical cues appeared to be related to the general risk category.

Table 2. A Critical Cue Inventory for Mental Health Professionals (including mean rankings).

Cue	Mean Rating	Category
Harm to self	4.80	Risk
Harm to others	4.80	Risk
Means	4.80	Risk
Intention	4.80	Risk
Support networks	4.57	Personal History
Conflicts in relationships	4.48	Personal History
Psychosocial	4.46	Personal History

Affect	4.43	Immediacy
Engagement	4.34	Immediacy
Emotional Expression	4.33	Immediacy
Coherence	4.32	Speech
Voice	4.27	Physical
Coping style	4.23	Immediacy
Communication style	4.21	Immediacy
Referral Question	4.20	Collateral
Lifestyle factors	4.18	Personal Info
Context appropriate	4.14	Speech
Developmental Appropriateness	4.12	Language
Relationship status	4.11	Personal History
Perceptions	4.10	Cognitive
Persaverative/Fixated	4.06	Speech
Divergent/Off topic	4.06	Speech
Personality/Temperament	4.03	Immediacy
Extent of insight	4.02	Collateral
Previous Diagnosis	4.01	Medical
Manner of narrative	4.00	Language
Attention	4.00	Cognitive
Decision-making	4.00	Cognitive

DISCUSSION

The aim of this study was to identify a critical cue inventory utilised by experienced MHPs for use in future training initiatives. This study sought to achieve this aim by: (a) using the critical decision method with a number of experienced MHPS to extract a range of cues used during the initial stages of psychological assessment; and, b) using a survey, investigate whether there are significant differences in MHPs ratings of perceived diagnosticity and frequency of use across the cues extracted, to determine the most critical cues.

The analysis yielded an array of cues; 73 across 11 general categories. This demonstrates that there are a range of cues present in MHPs' memory that can be matched, at least in part, to the operational environment during a client interaction, and which may be used by the MHP to guide the process of assessment. Rasmussen (1983) suggests that cues generate the recognition of critical conditions that may restrict the decision-maker searching for additional cues and their associations. In many ways, these cues appear to be predictive of certain outcomes such as disorders, and further, that these were used to guide the MHPs line of investigation for hypothesis generation, and further information seeking.

Participants' perceptions of cue diagnosticity and use were significantly correlated, with significant differences in perceptions for each pairing of cue, across both of these dimensions. This suggests that participants demonstrated a capacity to discriminate between fine gradations in the stimulus, a skill consistent with popular conceptions of expertise (Shanteau, 1991). Indeed, decision performance can be predicted by the percentage of relevant cues targeted by experts (Stokes et al., 1997) and experts attend to more relevant cues thereby improving their decision accuracy (Schriver et al., 2008). Based on this discrimination between cues in apparent relative value and use, the sample of cues was reduced to an inventory of 28 critical cues.

The findings underlines the relative importance of recognising a client's risk of harm to self and others. MHPs rated *Risk* cues as high in diagnosticity and frequency. The ratings of diagnosticity and frequency were substantially higher than those reported for any of the other cue categories identified. This is likely a reflection of the emphasis MHPs place on this information during professional training.

The Recognition Primed Decision model is regarded as combining both the intuitive and analytical processes of decision-making (Klein, 2008) and it appears that both these are evident in initial psychological assessment. Applying Loye (1983), Welsh and Lyons (2001), and Wittemann et al. (2012) definitions of intuition, firstly as the circumvention of linear cognitive processes, and secondly as the combination of formal knowledge stores and experience that results in automated response sets, it appears that MHPs rely both on practical experience and formalised knowledge stores accumulated during their academic training and continued professional development pursuits. All 12 of the interview participants referred to the application of theoretical models, principles and evidence-based treatments and instruments that they draw upon or referred to during the assessment decision task. Importantly, all of the interview participants reported that they required more than one session with a client to formulate an accurate mental representation, despite this however, they all indicated that there were important cues and associations identified in the initial assessment that formed the basis for the overall assessment decision task. This claim appears to be supported by the breadth and depth of the cues and cue categories elicited from the cognitive interview process.

Bhugra et al. (2011) emphasised experienced psychiatrists' reliance on intuition in guiding decisions, this may be an area that warrants further investigation to de-mystify the notion of intuition and perhaps incorporate intuitive clinical decision-making into MHP competent decision-making models and future training initiative developments. Wittemann et al. (2012) suggest that research should attempt to outline the intuitive processes rather just encourage MHPs to reflect upon intuitive decision-making retrospectively.

The recognition that the development of expertise takes time and effortful engagement within a domain is a luxury that may not always be available. It must be remembered that trainee MHPs generally engage with assess and treat clients within a supervised practicum framework near the finalisation of their academic training and this is usually well before they are considered expert in their practice. It seems reasonable therefore to promote the skill acquisition process and in this case, decision-making competency. This could be achieved through the use of cue-based training, delivered either by increasing trainee MHP's awareness of the cues available in the operational environment, or alternatively by incorporating the cues within simulated human-systems interface designs, such as a virtual patients. Simulation, as a form of cue-based training has been successfully applied within the aviation industry for testing and training pilot ability (Hays et al., 1992; Wiggins & OHare, 2003) and is probably used, at least in part, because there are features of the simulated environment that facilitate, believability, immersion, and presence (Glantz, Graap, & Rizzo, 2003).

Within the context of expertise, it should be noted that a core assumption of CTA is that the discovery of how decision tasks are performed necessitates the use of individuals proficient in the domain in order to generate content rather than process knowledge. That is, knowledge that can be modelled and learned by those considered less proficient in the domain. To date research applying CTA techniques has focused on experts and novices in domains such as aviation (Schriver et al, 2008), whitewater rafting (O'Hare et al., 1998), emergency control (Flin, Slaven, & Stewart, 1996), and criminal investigation (Morrison et al., 2013). The assessment decision task might be considered diverse from the aforementioned examples, partly because it is not always the immediate goal of the MHP to make critical decisions but rather to filter, synthesise and conceptualise substantial amounts of information presented by the client in verbal and non-verbal arrangements. Further, the current research did not aim to differentiate cue use across expertise, nor did it aim to elucidate the process knowledge of the decision task. Rather it aimed to explore the nature of MHP knowledge with a view to identify the content knowledge utilised, that is, the cues engaged during decision-making. Here it is assumed that MHPs utilise cues and their associations in the operational environment to create a synthesis to form an overall mental model of their clients.

Ostensibly, both the level of expertise and the nature of the operational environment may impact decision-making performance and Beilock et al. (2002) suggest that the absence of critical cues from the operational environment results in a reduction in the expert's decision-making accuracy. This may have implications for MHPs practicing in alternative operational technologically mediated environments such as the telephone, online email or Internet chat rooms. Psychologists have noted the increased use of technology in individuals' lives (Bee et al., 2008; Richards, 2009). As the reliance on technologically mediated methods for mental health service delivery increases and consistent with the general increase in technology use, this may imply that certain operational environments may impede or restrict the availability of critical cues. Indeed, the identification of those critical cues that are impacted due to the operational environment may be important both for the prospective adaptation or modification of alternative operational environments and for training provided to those MHPs practicing in the same. This point is underlined by the current findings, which reveal an emphasis on visual cues. Further, other more online methods of capturing the decision-making processes of MHPs (e.g., eye-tracking) should also be explored.

CONCULSION

Decisions in mental health assessment are complex, time bound, dynamic and often have important consequences. The NDM framework offers an opportunity to gain insight into the nature of decision-making in mental health assessment from an ecologically valid perspective, which facilitates the application of the results to inform future mental health training initiatives such as virtual patients.

This appears to be the first study to investigate decision-making processes of MHPs within the NDM framework. Klein (1993) offers that NDM is appropriate for use in naturalistic settings where the decisions are challenging in the context of time constraints, high stakes, unclear goals and dynamic conditions. Congruent with the aforementioned decision characteristics, mental health assessment can be considered information-rich, dynamic, often with serious implications or consequences and time-constraints. These characteristics of assessment decisions are particularly evident in the area of risk assessment where the conditions are dynamic, and in some instances the life of the client or others is potentially at risk.

The current research affirms that cue use is an important component of mental health assessment decision-making.These findings offer promising avenues for future research and may be an important first step in developing a greater understanding of the processes involved in MHP decision-making, particularly, those described as intuitive. Importantly, it is not the intention of this research study to reduce the process of MHP decision-making to a reductionist and mechanistic explanation, but rather it is an attempt to begin to unravel the complexities of MHP decision-making within a naturalistic and ecologically valid framework.

REFERENCES

Bee, P. E., Bower, P., Lovell, K., Gilbody, S., Richards, D., Gask, L., & Roach, P. (2008). Psychotherapy medicated by remote communication technologies: a meta-analytic review. *BMC Psychiatry, 8*(1), 60-69. doi:10.1186/1471-244X-8-60.

Beilock, S. L., Wierenga, S. A., & Carr, T. H. (2002). Expertise, attention, and memory in sensorimotor skill execution: Impact of novel task constraints on dual-task performance and episodic memory. *The Quarterly Journal of Experimental Psychology Section A Human Experimental Psychology, 55*(4), 1211-1240.

Bhugra, D., Easter, A., Mallaris, Y., & Gupta, S. (2011). Clinical decision making in psychiatry by psychiatrists. *Acta Psychiatrica Scandanavia, 124*, 403-411.

Broadbent, M., Moxham, L., & Dwyer, T. (2007). The development and use of mental health triage scales in Australia. *International Journal of Mental Health Nursing, 16*, 413–421.

Crandall. B., Klein, G. A., & Hoffman, R. R. (2006). *Working minds: a practitioner's guide to cognitive task analysis*. Cambridge, MA: The MIT Press.

Ebbesen, E. B., & Konecni, V. J. (1975). Decision making and information integration in the courts: The setting of bail. *Journal of Personality and Social Psychology, 32*(5), 805-821.

Fisher, D., & Pollatsek, A. (2007). Novice driver crashes: Failure to divide attention or failure to recognize risks. In A. F. Kramer, D. A. Wiegmann & A. Kirlik (Eds.), *Attention: From theory to practice* (Vol. 134-156). Oxford, UK: Oxford University Press.

Flin, R., Slaven, G., & Stewart, K. (1996). Emergency decision-making in the offshore oil and gas industry. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 38*(2), 262-277.

Glantz, K., Graap, K., & Rizzo, A. (2003). Virtual reality for psychotherapy: Current reality and future possiblities. *Psychotherapy: Theory/Research/Practice/Training, 40*(1/2), 1-13.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*, 19-30.

Hammond, K. R., Frederick, E. N., Robilllard, N., & Victor, D. (1989). Application of cognitive theory to the student-teach dialogue. In D. A. Evans & V. L. Patel (Eds.), *Cognitive science in medicine: Biomedical modelling* (pp. 173–210). Cambridge, MA: MIT Press.

Hays, R. T., Jacobs, J. W., Prince, C., & Salas, E. (1992). Requirements for future research in flight simulation training: Guidance based on a meta-analytic review. *International Journal of Aviation Psychology, 2*, 143–158.

Hershey, D. A., Walsh, D. A., Read, S. J., & Chulef, A. S. (1990). The effects of expertise on financial problem solving: Evidence for goal-directed, problem solving scripts. *Organizational Behavior and Decision Processes, 46*, 77-101.

Kenny, P., Parsons, T. D., Pataki, C., Pato, M., St-George, C., Sugar, J., & Rizzo, A. A. (2008). Virtual Justina: A PTSD virtual patient for clinical classroom training. *Annual Review of CyberTherapy and Telemedicine, 6*(1), 113-118.

King, L. (1997). Intuition: a critical review of research and rhetoric. *Journal of Advanced Nursing, 26*(1), 194-202.

Klein, G, A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fireground. In Proceedings of the Human Factors and Ergonomics Society 30th Annual Meeting (pp. 576–580). Santa Monica, CA: Human Factors and Ergonomics Society.

Klein, G. A., & Klinger, D. (1991). Naturalistic Decision Making. *Human Systems: IAC Gateway, 2*(1), 16-20.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision-making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), Decision making in action: Models and methods (pp.138–147). Norwood, NJ: Ablex.

Klein, G. A. (2008). Naturalistic Decision Making. Human Factors: *The Journal of the Human Factors and Ergonomics Society, 50*(3), 456-460. doi: 10.1518/001872008X288385.

Loveday, T., Wiggins, M. W., Searle, B. J., Festa, M., & Schell, D. (2012). The capability of static and dynamic features to distinguish competent from genuinely expert practitioners in pediatric diagnosis. *Human Factors: The Journal of Human Factors and Ergonomics Society, 55*(1), 125-137.

Morrison, B. W., Wiggins, M. W., Bond, N. W., & Tyler, M. D. (2013). Measuring relative cue strength as a means of validating an inventory of expert offender profiling cues. *Journal of Cognitive Engineering and Decision Making, 6*(4), doi: 10.1177/1555343412459192.

O'Hare, D., Wiggins, M., Williams, A., & Wong, W. (1998). Cognitive task analysis for decision centered design and training. *Ergonomics, 41*(11), 1698-1718.

Pachur, T., & Marinello, G. (2013). Expert intuitions: How to model the decision strategies of airport customs officers? *Acta Psychologica, 144*(1), 97-103.

Perry, N. C., Wiggins, M. W., Childs, M., & Fogarty, G. (2013). The application of reduced-processing decision support systems to facilitate the acquisition of decision-making skills. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 55*(3), 535-544.

Richards, D. (2009). Features and benefits of online counselling: Trinity College online mental health community. *British Journal Of Guidance & Counselling, 37*(3), 231-242. doi:10.1080/03069880902956975.

Schmidt, H. G., & Boshuizen, H. P. A. (1993). Acquiring expertise in medicine. *Educational Psychology Review, 3*, 205–221.

Schriver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise differences in attentional strategies related to pilot decision-making. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*(6), 864-878. doi: 10.1518/001872008X374974.

Shanteau, J. (1991). Psychological characteristics and strategies of experts. In G. Wright & F. Bolger (Eds.), *Expertise and Decision Support* (pp. 55-76). NY: Plenum.

Stokes, A. F., Kemper, K., & Kite, K. (1997). Aeronautical decision-making, cue recognition, and expertise under time pressure. *Naturalistic decision making* (pp p. 183-196), New Jersey: Lawrence Erlbaum Associates, Inc.

Vollmer, S., Spada, H., Caspar, F., & Burri, S. (2013). Expertise in clinical psychology. The effects of university training and practical experience on expertise in clinical psychology. *Frontiers in Psychology, 4*(14), 1-12.

Welsh, I., & Lyons, C. M. (2001). Evidence-based care and the case for intuition and tacit knowledge in clinical assessment and decision making in mental health nursing practice: an empirical contribution. *Journal of Psychiatric and Mental Health Nursing, 8*, 299-305.

Whaley, A. L., & Geller, P. A. (2007). Toward a cognitive process model of ethnic/racial biases in clinical judgement. *Review of General Psychology, 11*(1). 75-96.

Wiggins, M., & O'Hare, D. (2003). Expert and novice pilot perceptions of static in-flight images of weather. *International Journal Of Aviation Psychology, 13*(2), 173-187.

Wittemann, C. L. M., Spaanjaars, N. L., & Aarts, A. A. (2012). Clinical Intuition in mental health care: A discussion and focus groups. *Counselling Psychology Quarterly, 25*(1), 19-29.

The Challenges of Asynchronous Communication for Distributed Teamwork: Task Performance and Media Effects

Kathleen MOSIER[a], and Ute FISCHER[b]
[a] *San Francisco State University*
[b] *Georgia Institute of Technology*

ABSTRACT

The present study was conducted to examine how communication delay impacts distributed team performance, and whether communication media moderates or exacerbates its effect. Twenty-four teams of three collaborated remotely on computer-based tasks simulating failures in a spacecraft's life support system. Communication medium (text vs. voice) was a between-group variable; presence/absence of communication delay was a within-group variable. Transmission delay impacted time required to initiate a successful repair and more importantly, its effect varied by communication medium. When communication was delayed, teams used a comparable amount of time to repair system failures. However, when communication was synchronous, voice teams outperformed text groups. Likewise, teams' accuracy in performing system repairs was influenced by communication medium as teams communicating by text undertook more incorrect repairs than teams communicating by voice. The analysis of team communication identified differences between the text and voice teams that are consistent with medium-specific affordances and constraints.

KEYWORDS

Research; Macrocognition; Common Ground; Aerospace

INTRODUCTION

Effective and efficient communication between Mission Control and space crews is essential for successful task performance and mission safety. The importance of team communication is heightened when unforeseen problems arise, such as system failures that are time-critical and require extensive coordination and collaboration between space and ground crews. Examples abound from Apollo missions to the present day. Problems during dynamic phases of flight, such as the lightning strikes during ascent in Apollo 12 and the lunar module landing abort switch failure in Apollo 14, as well as problems with longer fuses but complex life-threatening implications as in Apollo 13, illustrate how critical interactions between flight crew and ground are in managing these complications - interactions that will be severely challenged with time delays. As missions travel further from the Earth, delays in communication will be unavoidable. During long duration missions and missions beyond Low Earth Orbit, space-ground communications will involve delays up to 20 minutes one way, a reality that poses a formidable challenge to team communication and ultimately to mission safety and success.

Investigations of asynchronous communication in domains such as telemedicine have identified communication delays as a primary impediment to effective telesurgery, and have prescribed faster transmission technology as the solution (e.g., Eadie, Seifalian, & Davidson, 2003). Given the current limitations of earth-space transmission technology, however, it is essential to explore solutions that focus on communication processes per se rather than transmission speed, and to devise process strategies to mitigate problems associated with asynchronous communication.

Common ground theory of communication (Clark, 1996), which emphasizes the interactive and goal-directed nature of communication and relating communication processes to constraints inherent in different communication media, served as the framework in the present research to examine the impact of delayed communication on remote team collaboration. Common ground theory views communication as a collaboration between speakers and addressees. Conversational partners need to coordinate the communication process (e.g., when to speak) as well as its content (e.g., speakers present information and addressees have to confirm their understanding or request clarification) to ensure that the information becomes part of their common ground—that is, is accepted as mutually understood, accurate and relevant to shared goals. To do so effectively, partners need to adapt their behavior to the opportunities and constraints associated with different communication situations and media (Brennan & Lockridge, 2006; Clark & Brennan, 1991; Olson, G. & Olson, J., 2007).

Generally it is more challenging to establish shared task and team knowledge when team members are spatially distributed than when they interact face-to-face as fewer resources are available. Communication with remote partners who are temporally co-present (synchronous communication, such as telephone or instant messaging) eliminates visual cues and thus requires more explicit communication; however, turn-taking with either voice or text can be rapid as messages can be received almost instantaneously, and their order easily determined. Voice communications maintain the aural meaning nuances of face-to-face interactions, which are lacking in text-based conversations. On the other hand, writing enables partners to re-read and thus remember past communications, and to review and revise their messages prior to sharing them with others.

Currently little is known about how communication delay will impact space-ground collaboration and task performance, and how different communication media may mediate its effect. The present study is part of a research program to address this knowledge gap. Specifically, we examined the effects of transmission delay on team communication, teamwork, and task performance under different media conditions. Communication medium and transmission delay were predicted to significantly impact team communication and task performance. Transmission delay was hypothesized to be associated with decrements in task performance, and to disrupt the coherence of team communication making it difficult for distributed team members to establish common ground. These effects were predicted be most evident when team members relied on voice communication.

METHOD

Design

The design was a 2 (communication medium – voice vs. text) x 2 (task type – simple vs. difficult/problem-solving AutoCAMS failure) x 2 (time delay – no delay vs. 5-min delay) mixed design. Communication medium was a between-teams variable; task type and time delay were varied within teams.

Participants

The study included 72 (24 teams of 3) undergraduate and graduate students between the ages of 21-55. All participants were fluent English speakers, had at least two years of college, and had experience with computers.

Experimental Task

The micro-world for this study, AutoCAMS 2.0 (Manzey, Bleil, Bahner-Heyne, Klostermann, et al., 2008), simulates the life support system of a spacecraft and requires team members to monitor and control different subsystems. This micro-world mimics critical aspects of flight crew activities during space operations and has considerable face validity. The interface (see Fig. 1) contains real time and trend information for 5 subsystems, carbon dioxide, oxygen, pressure, temperature, and humidity and includes an Automatic Fault Identification and Recovery Agent, which can be programmed to give true, false, or ambiguous indications of system failures.

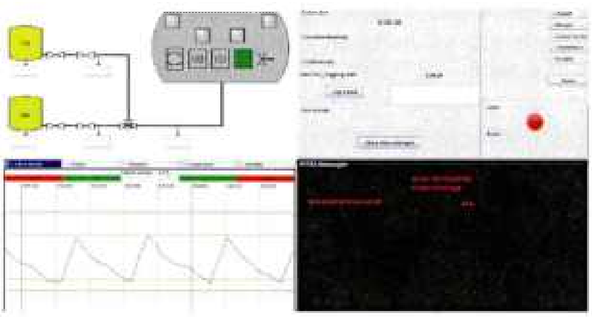


Figure 1. AutoCams display

Procedure

Teams of three participants were randomly assigned to one communication mode condition. One participant in each team was assigned to the role of Flight Systems Engineer (FSE) and AutoCAMS expert onboard the fictional US Space Station; the other two participants were told they were *Pioneer* spacecraft crewmembers on an exploratory mission in deep space. In order to guarantee the requirement of communication and collaboration on the experimental tasks, task-related expertise concerning diagnostic and repair procedures were differentially distributed among team members. Team members were given 2-4 hours training, depending on their roles. The FSEs received extensive training on AutoCAMS systems, diagnoses, and repairs, and had access to a comprehensive reference manual. Pioneer crewmembers were given basic training on AutoCAMS: they were trained on how to access diagrams of its systems, but did not receive any instruction on failure diagnosis and repair. When a failure occurred on their system, the *Pioneer* crew had to contact the FSE for guidance on the diagnostic process and repair procedures. The collaborative demands of this situation were analogous to events in space operations for which astronauts lack in-depth expertise; for example, the medical emergency situation included a space-analog simulation study conducted by Frank and colleagues (Frank, Spirkovska, McCann, Lui, et al. 2013). Each role also required several secondary tasks that tapped prospective memory, reaction time, and attention to detail, and were intended to provide a moderate-to-high level of workload, similar to the workload astronauts might experience. FSE-Pioneer teams collaborated in two 90-min flight segments for which time delay and task order were counterbalanced.

Task type. During each flight segment, the Pioneer spacecraft experienced two AutoCAMS malfunctions for which the crew needed assistance from the Flight Systems Engineer. One failure was simple (the automated alarm specified the failure) and its diagnosis involved only confirmation of diagnostic parameters before repair. The more complex failure presented the crew with ambiguous system indications and required several back-and-forth communications with the FSE to discover the specifics of the malfunction and prescribe the appropriate repair.

Communication mode. The Pioneer crew communicated with the FSE via either text or voice as assigned. Text-based communications used Pidgin, a multiprotocol instant messaging program. Voice communications between Pioneer crews and the FSE employed a voice over internet protocol (VOIP). Transmission delay of voice and text communications in the asynchronous flight segment was achieved by routing them through a Linux-based emulator developed by NASA engineers at Kennedy Space Center and set up on a SFSU server.

Task performance measures. Task performance data (i.e., interactions with the AutoCAMS system) were collected by the computer-based experimental task. Task performance was measured in terms of time required to initiate a successful repair as well as accuracy of the repair procedure. The duration of a failure was measured from the appearance of a red alarm and corresponding failure message to the initiation of the correct repair. Performance data on secondary tasks were also collected but will not discussed here.

Communication measures. Communication analysis focused on the interactions between the FSE and the Pioneer crew during the failure repair tasks. Audio-recordings of the voice communications between the crewmembers and the FSE were loaded into Audacity audio-editing software and transcribed for subsequent analysis. Logs of team members' text-based communications were directly uploaded for analysis.

The unit of analysis for the communication coding was a turn. In the voice condition, "turn" refers to an uninterrupted speech segment by a speaker usually marked by turn signals (e.g., *Thanks*; or *Ok*?), falling intonation, or pauses. In text-based communications, any text written by a participant before pressing the send button constitutes a turn.

Communication analyses examined quantitative characteristics (i.e., communication rate and length) as well as structural aspects and content variables. Structural aspects concerned *information splitting* (related information, such as diagnostic cues, is presented in separate turns) and the *distance between adjacency pairs* (i.e., the number of turns intervening between pairs of related communications by conversational partners, such as question-answer). Content coding focused on communication problems, egocentricity, and threats to common ground as well as strategies aimed at managing communication delay. *Communication problems* include instances in which a team member asked a partner to clarify information or repeated an earlier contribution that had not been acknowledged by the partner. *Egocentricity* refers to instances in which a team member displayed adjacency bias (i.e., misinterpreted a partner's contribution that immediately followed his/her own most recent communication as a response to it), or insensitivity to the transmission delay (i.e., conversational partner repeated information or requested feedback before he/she could have received a response). *Threats to common ground* included missing responses (i.e., failure of an addressee to respond to or acknowledge contributions by his/her partner), responses that provided incorrect or incomplete information or were out of sequence (i.e., response was received after events had rendered it outdated), and the *use of terms whose meaning was underspecified* (e.g., "We have a problem") or could not be established within a turn but rested on information in preceding turns (for instance, "We completed the repair" required the addressee to remember elements in previous communications to identify the relevant repair). The coding of communication strategies focused on efforts by team members to mitigate the disruption of the turn sequence and the cognitive demands posed by communication delays. The following strategies were discerned: *Indicate end of one's turn* (e.g., by saying "Bye," or "Out"); give partner a *heads-up* (i.e., alert partner to upcoming transmission of critical information); prefix own message with *topic* or *refer to a partner's preceding contribution*; present *complex information in* a *structured fashion* (e.g., by numbering steps in repair procedure); *highlight critical information* (e.g., by repeating it within the same turn); and *push information* (i.e., volunteer critical information before partner requests it).

RESULTS AND DISCUSSION

Discussion of findings will focus on team communication and performance under time-delayed conditions and in relation to different communication media. Performance and communication data of teams as they communicated synchronously will be provided but not discussed in detail.

Task Performance

Mixed-design Analysis of Variance (ANOVAs) on time in red indicated that as predicted teams took significantly longer to repair system failures under time delay (TD) than when they had no time delay (NTD). $F(1,22)=7.54$, $p=.012$, partial $eta^2=.253$. Surprisingly, this difference was concentrated only in the voice medium, as reflected in a significant time delay x medium interaction, $F(1,22)=7.98$, $p=.01$, partial $eta^2=.266$. Under time delay, teams using either media performed comparably in terms of time to repair. When communications were synchronous, however, the voice condition provided an advantage and voice teams took significantly less time than text teams on system failure tasks (see Table 1). No significant effects of medium ($F(1,20) = .001$, ns.) or transmission delay ($F(1,20) = 2.67$, ns.) were observed on number of correct repairs. Two crews attempted an unusually high number of repairs without direction from the FSE, and were excluded as outliers from any analyses of the number of correct and failed repairs.

Table 1. Task performance measures

	Time in Red (in min) (N=24)		Correct Repairs (N = 22)		Incorrect Repairs (N = 22)	
	TD	NTD	TD	NTD	TD	NTD
TEXT	56.21 (24.99)	56.76 (20.54)	1.40 (.8433)	1.70 (.6750)	3.4 (2.76)	5.8 (3.74)
VOICE	61.63 (20.77)	29.84 (25.65)	1.42 (.7930)	1.67 (.7785)	1.58 (2.02)	2.75 (2.80)

The number of incorrect repairs that Pioneer crews initiated was also analyzed as a measure of performance. Significantly more incorrect repairs were committed in the text condition than when communicating via voice, $F(1,20)=10.16$, $p=.005$, partial $eta^2=.149$; though this effect seems mainly driven by the NTD condition. The data suggest that the NTD condition may have been more conducive to incorrect repairs than the TD condition, $F(1,20)=3.50$, $p=.076$, partial $eta^2=.337$. Both of the outlier crews excluded from these analyses were text medium crews, and each crew made an excessive number of failed repair attempts during the NTD leg, providing additional support for this interpretation.

Team Communication

The presence of transmission delay impacted the structure of team members' communications. Separate Analyses of Variance on communication rate and density revealed that communication delay influenced both the rate of turns by team members ($F(1,20)=87.80$, $p<.0001$, $eta^2=.81$) and the length of their contributions ($F(1,20)=74.36$, $p<.0001$, $eta^2=.79$)[9]. As can be seen in Table 2, team members made fewer but longer contributions when they communicated under time delay than when no time delay was present, irrespective of the communication medium they used. Moreover, as shown in Table 2, these effects were more pronounced for teams communicating by voice than those communicating via text. The medium by time delay interaction was significant for rate of communication ($F(1,20)=26.39$, $p<.0001$, $eta^2=.57$) as well as for the length of the communication ($F(1,20)=42.63$, $P<.0001$, $eta^2=.68$). In response to the transmission delay, voice teams reduced their communication rate by a factor of 13; text teams' rate decreased by a factor of 3. Interestingly, while voice and text teams communicated in comparable rates under delayed conditions, team members in the voice condition made contributions that were considerably longer (Mean number of words/turn = 61.84) than communications by text teams (M = 13.5). This finding suggests that team members using text may have been more concise than team members in the voice condition. However, as subsequent content analyses indicate text communication was also associated with an increased potential for misunderstanding.

Table 2. Communication rate and density in text and voice teams during TD and NTD conditions

	TEXT		VOICE	
	NTD	TD	NTD	TD
Communication rate (turns/min)	2.5	.86	6	.46
Communication density (words/turn)	6.43 (1.63)	13.5 (6.14)	10.73 (4.08)	61.84

Content analysis of team communication focused on Pioneer Crew/FSE interactions during transmission delay. Medium-specific differences concerned structural aspects of team communication as well as content variables. As can be seen in Table 3, text teams were more likely than voice teams to split up related information and present it in separate turns ($F(1,22)=15.48$, $p=.001$, $eta^2=.41$) and to have more turns come between related communications (adjacency pairs such as question and answer) by distributed team members ($F(1,22)=7.03$, $p=.015$, $eta^2=.24$). Text team members' communication also showed more instances of communication problems ($F(1,22)=5.68$, $p=.03$, $eta^2=.21$) where a team member indicated non-understanding and requested clarification, or repeated his/her contribution after a partner failed to provide feedback. Text communication also included more threats to common ground ($F(1,22)=7.24$, $p=.01$, $eta^2=.25$), in particular missing responses ($M_{Text} = 5.08$; $M_{Voice}=2.17$) and ambiguous terms (i.e., terms whose meaning was underspecified, or could not be established within a turn but rested on information in preceding turns; $M_{Text} = 19.42$; $M_{Voice}=8.58$).

These differences are consistent with medium-specific affordances and constraints. Text provides team members with a written record of their on-going conversation, and thus may enable them to keep track of related contributions and the identity of referents across turns. However, as the presence of communication problems in the text group indicates team members may have overestimated the benefits of text-based communication. Voice communication is cognitively more taxing than text-based communication insofar as participants need to remember their ongoing discourse to interpret new information. Voice teams apparently adapted to this constraint by packing more information into one turn than text teams, behavior that kept related communications more closely aligned and may have aided comprehension.

Table 3. Structure and content of communications by text and voice teams during TD

	TEXT	VOICE
Turns intervening between adjacency pairs	10.64 (4.59)	6.57 (2.66)
Information splitting (% of turns)	16.51 (9.33)	3.43 (6.80)
Communication problems	2.67 (1.23)	1.17 (1.80)
Threats to Common Ground	31.33 (17.48)	15.67 (10.06)
Egocentricity	4.75 (3.75)	3.75 (3.28)

Further analyses examined the communication strategies three high- and low-performing teams in each medium condition employed to identify measures supporting team collaboration under time delay. Mean times (in min) to repair system failures for these teams were: $Text_{High} = 32.2$; $Text_{Low} = 84.35$; $Voice_{High} = 33.83$; $Voice_{Low} = 83.57$) Table 4 shows that high-performing teams, in particular high voice team members, relied on several strategies that may have helped them to maintain conversational coherence when communication was delayed. They identified messages by topic, presented information in a well-structured manner and repeated critical information, apparently in an attempt to facilitate comprehension.

Members of high-performing voice teams also seemed attuned to the fact that their perspective on evolving events may be different from their remote partners as a result of the time-delay. They tended to push information to remote partners in a timely manner.

Table 4. Communication strategies used by high- and low-performing text and voice teams during TD

	TEXT		VOICE	
	High	**Low**	**High**	**Low**
Indicate end of turn	0	0	29.89 (35.25)	24.14 (30.17)
Heads-up	0	0	0	5 (6.08)
Provide topic or reference to previous turn	10.17 (10.43)	1.67 (2.89)	22.71 (12.52)	12.14 (5.92)
Structure/chunk complex information	11.45 (10.57)	5 (8.66)	9.70 (10.01)	6.96 (7.45)
Highlight critical information	0	0	8.02 (9.28)	6.96 (9.40)
Reference time	0	0	12.12 (20.99)	17.71 (17.74)
Push information	5.24 (6.49)	2.57 (2.32)	20.27 (9.09)	8.27 (11.63)

Note. Numbers indicate mean percentage of turns during TD flight segment that adhere to a given strategy; SD in parenthesis

However, in both text and voice teams instances of miscommunication in which team members failed to account for the communication delay were evident. Team members displayed adjacency bias; that is, they mistook a remote partner's communication that immediately followed their own transmission as a response to it, or they showed insensitivity to the delay by repeating a message before they could have received a response from their partner. These instances required additional communication in which team members clarified their situation understanding, or they spiraled into misunderstanding from which team members never recovered and thus were unable to repair a system failure. This situation is depicted in Figure 2 that summarizes the first 20 minutes of dialogue between a Flight Systems Engineer (FSE) and his Pioneer crew (P) after the occurrence of a system failure. The lower portion of the graph presents the turn sequence from P's perspective; the top portion shows the temporal sequence of the same turns as experienced by the FSE. Colored rectangles represent individual turns by P or FSE; numbers (e.g. P1, P2 ..., F1, F2 ...) designate the first, second etc. turn by a Pioneer crewmember or the Flight Systems Engineer. As can be seen the temporal sequence of turns differs for P and FSE, and more importantly, contributions that for one partner are related and following each other (as the FSE's response (F2) to P1) are not adjacent for the other partner. If team members misalign contributions, serious misunderstandings can arise. This problem happened to the crew whose discourse is depicted in Figure 2. The Pioneer crew erroneously assumed that the repair instruction provided by the FSE in F8 was a response to their failure announcement in P17. However, the FSE's instruction was in response to previous requests by the Pioneer crew (P9 and P10) and was incompatible with their current malfunction. The team never recovered from this misunderstanding and failed to repair their second system failure.

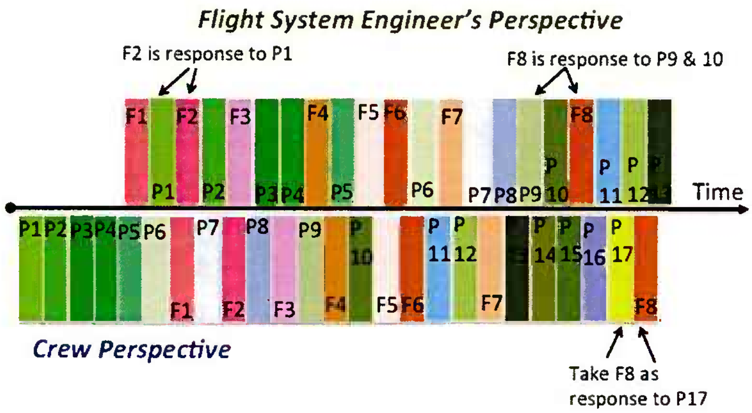


Figure 2. Representation of the same conversation under TD as perceived by different team members

CONCLUSIONS

Perhaps the most interesting finding of the present study was that when communication was delayed by 5 minutes, task performance of distributed teams was comparable irrespective of the communication medium they used for collaboration. That is, neither of the communication media we investigated –voice or text—was better suited for remote collaborations under time-delayed conditions. This finding was not as predicted since we had hypothesized that text communication would provide an advantage over voice communication. On the other hand, consistent with our predictions we observed that transmission delay disrupted the turn sequence of remote partners' contributions and led to misunderstandings. Our research further suggests that successful teams in each media condition were those who adapted to the constraints of their communication medium to establish shared task understanding. These insights have informed the design of media-specific communication protocols in support of mission control–space crew communication and collaboration under time-delayed conditions. The effectiveness of the communication protocols is currently being assessed in laboratory research as well as in several space analog studies.

The present study also demonstrates the validity of conducting research involving a student population. While communication problems were more pronounced in the student population than in the research involving astronauts (Fischer, Mosier & Orasanu, 2013; Palinkas, 2013) –for instance with respect to the use of ambiguous terms—the nature of the problems was identical. Specifically, in both populations we observed adjacency bias and insensitivity to transmission delay, as illustrated by the following quote of a NEEMO 16 participant: "We looked at the voice loops, we looked at the text loops that occurred during these scenarios, and we saw afterwards that it was broken ten ways to Sunday. We were talking past each other; we were taking one response to mean, to be a response to a totally different question, you know, it was incredibly broken, and you could only see it when you took the time to really analyze it afterwards" (quoted in Palinkas, 2013). Likewise, the student participants in the present study generated the same strategies as astronauts (Fischer, Mosier & Orasanu, 2013).

References

Brennan, S. E., & Lockridge, C. B. (2006). Computer-mediated communication: A cognitive science approach. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics, 2nd Edition (pp. 775-780).* Oxford, UK: Elsevier Ltd.

Clark, H. H. (1996). *Using Language*. Cambridge, UK: Cambridge University Press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127 –149). Washington, DC: American Psychological Association.

Eadie, L. , Seifalian, A. , & Davidson, B. (2003). Telemedicine in Surgery. *The British Journal of Surgery, 90*(6), 647-658.

Fischer, U., Mosier, K., & Orasanu, J. (2013). The impact of transmission delays on Mission Control – Space Crew communication. In *Proceedings of the Human Factors and Ergonomics Society 57th Annual Meeting* (pp. 1372-1376). Santa Monica, CA: HFES.

Frank, J., Spirkovska, L., McCann, R., Lui, W. et al., (2013). Autonomous Mission Operations. IEEE Aerospace Conference, 1-20.

Krauss, R., & Bricker, P. (1967). Effects of transmission delay and access delay on the efficiency of verbal communication. *Journal of the Acoustical Society of America, 41*, 286-292.

Kraut, R. E., Fussell, S., R., Brennan, S. E., & Siegel, J. (2002). Understanding the effects of proximity on collaboration: Implications for technologies to support remote collaborative work. In P. Hinds & S. Kiesler (eds.), *Distributed work* (pp. 137-162). Cambridge, Mass: MIT Press.

Love, S. G. & Reagan, M. L. (2013). Delayed voice communication. *Acta Astronautica, 91*, 89-95.

Manzey, D., Bleil, M., Bahner-Heyne, J. E., Klostermann, A., Onnasch, L., Reichenbach, J., & Röttger, S. (2008). *AutoCAMS 2.0 Manual.* Berlin: Technical University of Berlin.

Olson, G. M., & Olson, J. S. (2007). Computer-supported cooperative work. In F. T. Durso, R. S. Nickerson, S. T. Dumais, S. Lewandowsky, & T. J. Perfect (eds.), *Handbook of Applied Cognition* (pp. 497-526). Chichester, England: Wiley.

Palinkas, L. (2013, Feb. 11-14). *The impact of communication delay on individual and team performance.* Poster presented at BHP Investigator meeting, Houston, TX.

ACKNOWLEDGMENTS

Funding for this research was provided by Cooperative Agreement NNX12AR19G between NASA Johnson Space Center, Georgia Institute of Technology and San Francisco State University. We are grateful to the following students at SFSU: Laura Carucci, Danielle Fox, Kathy Gonzalez, Jessica Lam, Daniel Maurath, Alec Munc, Kendra Reich, James Swarts, and Savanna Valdes. Their professionalism and dedication made this research possible. We also thank Sandeep Manchem for his assistance with AutoCAMS.

E-Kardex: Observing the use of Sharp-End Generated 'Brains' for Informing the Design of a Hybrid System

Austin F. MOUNT-CAMPBELL[a], Esther CHIPPS[b,c], Valerie MOORE[c], , David D. WOODS[d], Kevin EVANS[a], Emily S. PATTERSON[a]

[a]*School of Health and Rehabilitation Sciences, The Ohio State University*

[b]*College of Nursing, The Ohio State University*

[c]*Nursing Informatics, The Ohio State University Wexner Medical Center*

[d]*Integrated Systems Engineering, The Ohio State University*

ABSTRACT

Nurses are considered to be one the most adaptive and resilience producing clinician types; they are also one of the biggest users of paper based cognitive artifacts. Cognitive artifacts display information in an external manner to support cognitive processing and recall. The starting assumption is that there is a gap between nurses work-as-done and work-as-imagined such that nurses are generating cognitive artifacts in order to bridge the gap. In order to understand why nurses generate their own cognitive artifacts, the functional usage and value needs to be investigated. This study is anticipated to inform design requirements for EHR-generated cognitive support artifacts which are printed as a handoff report tool at the beginning of a nurse's work shift.

KEYWORDS

Research; Artifact; Healthcare; Activity management; Electronic health record; Hybrid system

INTRODUCTION

Cognitive artifacts display information in an external manner to support cognitive processing and recall (i.e., augment "knowledge in the head"). In the healthcare domain, nurses extensively rely on cognitive artifacts which display patient information, including lab results, radiology images, allergy lists, clinical flowsheets, vital signs, etc. Two primary electronic artifacts used extensively by registered nurses are electronic health records (EHRs) and electronic medical administration records (e-MAR). The stated intent of these artifacts at the most advanced level of implementation (Stage 7 in the HIMSS adoption model) is to replace the need for the use of paper as cognitive artifacts in the hospital setting (i.e., to have a 'paperless hospital'). These artifacts are required to be used by the 'blunt end', meaning that they are formal tools provided by administrators. Nevertheless, across all known hospitals, the majority of nurses reminisce fondly about functionality previously afforded by paper-based Kardex systems with individual patient summaries kept at a nursing station and maintained across shift boundaries and the use of paper to augment these artifacts is omnipresent with "brains" personal information sheets personally created by individual nurses at the beginning of their shifts. Therefore, we believe that identifying the functionality and value of these two cognitive artifacts will yield 'cognitive gold' in the sense of providing insights for designing a hybrid system that combines electronic and paper resources to support nurses' critical thinking, plan development, care delivery, and remembering elements to include in communications with other clinicians and during a shift change handover.

The conceptual lenses used in the proposed study are heavily influenced by Woods & Cook's (2010) conceptual framework for how organizational 'blunt end' factors shape adaptations of expert actors to use knowledge embedded in cognitive artifacts to meet goals despite environmental obstacles in evolving situations. This framework is displayed in Figure 1 and contains evolutions of elements from Neisser's seminal perception-action cycle (Neisser, 1976). This framework modifies schema by environmental exploration based on actors' knowledge, mindset and goals. While our focus is on the 'sharp end' nurses in this study who are providing direct patient care, the goals at the organizational level, i.e. 'blunt end', are ever-present.

Literature review: What Nurses are creating as a Cognitive Artifact

Nurses are considered to be one the most adaptive and resilience producing clinician types; they are also one of the biggest users of paper based cognitive artifacts (Gurses, Xiao, 2006). A literature review was conducted which focused on nurses and their interaction with personally created (sharp end) or organizationally implemented (blunt end) cognitive artifacts. Typically, sharp end generated cognitive artifacts are characterized as workarounds in that they are either not actively supported, or even actively discouraged, by the blunt end. Workarounds are defined as a deviation from an intended work process (Lowry, et al. 2015). An area for improving patient safety is reducing the gap between work-as-imagined (typically documented in policies and procedures by the blunt end) and work-as-practiced (typically based on direct observations of the sharp end). The starting assumption is that there is a gap between nurses and the sharp end or nurses would not be generating cognitive artifacts in order to bridge the gap. Overall, workarounds can be both positive and negative. Positive workarounds tend to be unexpected uses for features that were designed for a different purpose, and are typically performed first by individuals and then spread through personal networks, infrequently being spread to all people in a particular role. Negative workarounds tend to be unsafe and improve efficiency or the quality of work life at the expense of safety. There are different reasons for workarounds. There are workarounds which are required because the system does not allow work to be done as imagined, which are done to improve efficiency while increasing safety risks, which are done because of misaligned organizational incentives, and which are sub-optimal and are done because of failing to move to better processes for a variety of reasons. Each of these different types of workarounds has different implications for how to redesign systems and processes. The nurse developed brain is a general workaround performed by individual nurses. In the proposed dissertation study, the brains will be characterized using workarounds as part of the conceptual framework. They act as an example of work-as-practiced, where the brains are a positive workaround that are self-paced and in preparation for event driven activities such that the brains can provide a quick reference to aid recall when time is valuable. The brains are developed for personal use and are not part of the patient's official medical record.

<u>Figure 1. Conceptual framework diagram (Cook and Woods, 2010).</u>

In order to understand why nurses generate their own cognitive artifacts, the functional usage and value needs to be investigated. Typically, sharp end generated artifacts that supplement EHRs and e-

MARs do not contain unique information; in other words, the data are theoretically available in the existing blunt end provided artifacts. Table 1 provides a summary list of functional usage and value for cognitive artifacts, which is extended and modified from a framework created by McLane and colleagues (2012).

Table 1. 12-Typical Baseline Functions for Nursing Artifacts (expanded from McLane et al, 2012)

Functional Usage	Value
Providing a quickly accessible standardized location for finding key information	Develops a snap shot of clinical problems, current condition and care needs; acts as a quick reference
Aiding cognitive processing and internal memory storage	Handwriting improve memory recall of data
Organization of related information in spatial proximity and chronologically	Generates visual cues in order to refocus and get back on track after interruptions
Visual cues to highlight against a background	Highlight important information that is likely to be important
Balance workload and allocate resources	Allows nurses to manage their patient needs with the workload limits for themselves and their clinical collaborators
Provide reference for information needed to access additional information in electronic artifacts and from others	Provide identifying information for patients and contact information for specialist providers
Add, remove, and modify a list of "to do" action items	Develop a guideline or schedule for care planning needs throughout the shift
Support interacting with patient for medication reconciliation task with detailed ordered information	Acts as a reference verification of medication required and when
Chart data over time	Allows nurses to see vital sign changes over time
Support shift change handover	Provides nurses with a quick patient reference and memory aid for verbal updates during handovers
Cross-disciplinary communication	Provides a way to acquire or point to relevant patient information that can be utilized with ad hoc opportunistic interdisciplinary communication conducted away from a computer location

When nurses are asked why they use artifacts in favor of computers the answers typically fall into one of four categories: 1) Computer log-in and updating structured documentation is inefficient and time consuming, 2) Much of the information on the artifact does not have sufficient value over the long-term to be included in the formal chart, 3) Although the information is theoretically available in the EHR and/or e-MAR, it is organized poorly and it is difficult to locate it, and 4) The paper medium has advantageous elements compared to the electronic medium, such as being able to see what information was recently added and what information was previously available before an update (Gurses, Xiao, 2006).

Pilot Interviews: The Kardex: Why do nurses get excited?

Two pilot interviews were conducted with nurses that provide preliminary insights about the functional usage and value of the 'old' Kardex paper-based system, which is pictured in Figure 2 next to newer versions which are electronic or hybrid systems. The main insights were that the Kardex was value in that

"It was a compact, mobile, organized structured system, where you always knew where the information was…it was like an operations manual"

"Its like a guidebook for nurses"

It was small (5 in X 8 in), , one card per patient, and organized by bed location

In contained a summary of patient information, including important historical events

It supported shift change handovers for both bedside nurses (who listened as a group to all updates) and charge nurses

The interviews also identified issues with the old Kardex system, including:

The medication Kardex included information for all the patients in one place, and tended to be updated less frequently than the other portions of the system

The Kardex system did not support planning care activities or coordinating activities between registered nurses and personal care assistants

The nurses frequently had personal sheets of paper containing notes which were updated throughout the shift and discarded at the end (which likely were an early and less complex version of "brains" than are used currently)The generaly feeling during interviews is nurses loved the Kardex because of how the information was presented it was a one stop shop for their patients information, and it was easy to access contained within a single card. The issues with the Kardex was medication lists were not always up to date, and patient information for patients with long length of stays often ran out of room on the kardex. It was hypothesized by the nurses that the Kardex disapeared because the nursing job tasks evolved and the sharp end required use of other cognitive artifacts, most recently the EHR.

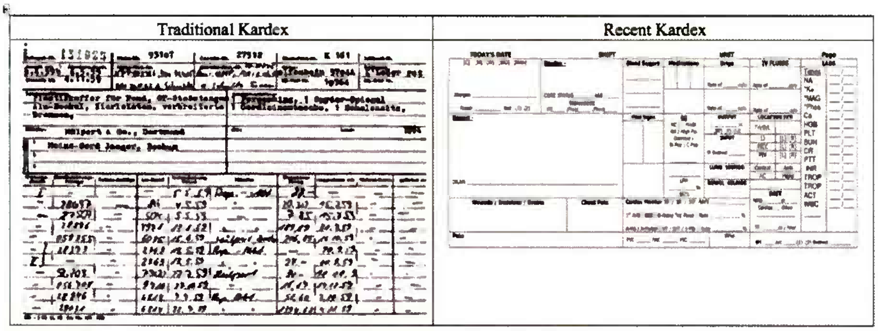


Figure 2 Examples of Traditional and Recent Kardex System

PROPOSED STUDY

Research Design

The design is a mixed-method study with the following components:

Semi-structured interviews of nurses about the functional usage and value of the traditional paper-based Kardex,

Ethnographic observations of registered nurses foraging in the EHR, generating the brains, utilizing it during the initial patient assessment process, and using it during shift change handover

Ethnographic observations of access and functional usage of cognitive artifacts (EHR components, e-MAR, and brains)

Digital photographs of "brains" will be taken during the ethnographic observations at the conclusion of the prior shift, after the new brains are generated, after all patients under a nurses' care have been assessed, and at the end of the four hour observation period

Sample

The sample is:

Semi-structured interviews: 40 registered nurses with Kardex experience obtained as a convenience sample from personal networks

Anticipated interview questions include:
What is your current nursing role?
What do you think are the greatest challenges related to documenting in the EHR for nurses?
Talk about the Kardex system you used to use. Advantages? Likes? Dislikes?
Was the Kardex supplemented with brains?
What information do you handwrite on an artifact that can be found in the EHR? Not in the EHR?
How does the structure and order of your personal artifacts differ from the EHR? How do you organize your artifact and why?
What else do you think we should be asking about Kardex use? Brain use?Who else do you think that we should talk with about this?

Ethographic observations: Two observation periods for each of 18 nurses during a four-hour period from the beginning of the work shift, including the prior handover, generating the brains, and initial assessment of all patients under their care. Nurses are evenly divided across four participating acute care units from two hospitals in a large academic medical center.

De-identified handwritten notes will be written on spiral paper about strategies and challenges to documenting in the EHR during the observations. Characteristics of brains sheet will be tabulated. Specifically, we will measure the number of categories of items documented on the sheet at the beginning, the number of categories of items added while seeing the patients the first time, and the number of categories of items added during the first four hours of the shift. Categories will be based upon grounded theory bottom-up analysis of the content of the sheets, but are likely to include identifiers, diagnoses, labs, procedures, action items, and contact information. The constant comparison method, a component of grounded theory, will be used to identify strategies. A codebook will be developed iteratively for strategies. Independent coders will classify strategies. An inter-rater reliability kappa score above 0.70 will be used to determine sufficient reliability across coders. Differences in codes will be resolved by discussion.

DISCUSSION

This study is anticipated to inform design requirements for EHR-generated cognitive support artifacts which are printed as a handoff report tool at the beginning of a nurse's work shift. We are anticipating findings that support elements such as a hybrid electronic/paper system, format suggestions for selected information on an overview summary single printed page, strategies for minimizing documentation time to update electronic structured data, and support for conducting handovers using a blunt end-designed protocol for ordering content

ACKNOWLEDGMENTS

Funding was provided by the OSUMC Patient Safety Advancement Grant Initiative (PI Patterson). The views do not necessarily represent the views of the OSUMC.

REFERENCES

Cook, R. I., & Woods, D. D. Chapter 3: Operating at the Sharp End. In *Behind Human Error*. 2010. Ashgate Publishing Company.

Gurses AP, Xiao Y. A systematic review of the literature on multidisciplinary rounds to design information technology. JAMIA 2006;13(3):267–76.

Klein, G. (2007). Flexecution, Part 2: Understanding and Supporting Flexible Execution: IEEE Intelligent Systems, 22, 108-112.

Lowry, S. Z., Ramaiah, M., Patterson, E. S., Brick, D., Gurses, A. P., Ozok, A., ... & Gibbons, M. C. (2014, June). Integrating Electronic Health Records into Clinical Workflow An Application of Human Factors Modeling Methods to Ambulatory Care. In *Proceedings of the International Symposium of Human Factors and Ergonomics in Healthcare* (Vol. 3, No. 1, pp. 170-177). SAGE Publications.

McLane, Sharon and James P Turley. "One Size Does Not Fit All: EHR Clinical Summary Design Requirements for Nurses." Nurs Inform (2012): 283-287.

Neisser, U. (1976). *Cognition and reality.* principles and implications of cognitive psychology (p. 230). W H Freeman & Co.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological review*, *84*(3), 231.

Patterson, E.S., Rogers, M.L., Chapman, R.J., Render, M.L. (2006). Compliance with Intended Use of Bar Code Medication Administration in Acute and Long-Term Care: An Observational Study. Human Factors, Special Issue on Patient Safety 48(1), 15-22.

Woods, D.D., & Patterson, E.S. (2001). How unexpected events produce an escalation of cognitive and coordinative demands. P.A. Hancock and P.A. Desmond (Eds.). Stress Workload and Fatigue. Lawrence Erlbaum, Hillsdale, NJ, pp. 290-304.

Look Who's Talking - In-game Communications Analysis as an Indicator of Recognition Primed Decision Making in Elite Australian Rules Football Umpires

Timothy J. NEVILLE[a,b], Paul M. SALMON[a]

[a] *University of the Sunshine Coast Accident Research, The University of the Sunshine Coast, Sippy Downs, Queensland, Australia*

[b] *Joint and Operations Analysis Division, The Defence Science and Technology Organisation, Canberra, Australia*

ABSTRACT

Officials in sport operate in a naturalistic environment making rapid decisions under stress. In sport, decision making research has identified consistent results between the three different 'variations' of the Recognition Primed Decision (RPD) model. This paper presents the findings from a study applying the RPD model to the decision making of Australian Rules Football (AFL) umpires. **Method:** Audible communication instances from AFL Field umpires were transcribed. The data was coded into 'decision moments'; each decision moment was analysed to identify if the decision conformed to one of the three RPD model variations. **Results:** Within the 6025 communication instances 887 decision moments were identified. 78% of the decision moments were classified as Variation 1, 18% as Variation 2 and 3.5 % as Variation 3. **Discussion:** Decision making in AFL umpires is characterized by a similar RPD breakdown as decision making by players in sport. AFL umpires RPD variation is influenced by the game situation and type of adjudication being made.

KEYWORDS

Judgement and Decision Making, Expertise, Officials in Sport, Australian Rules Football, Communications Technology,

INTRODUCTION

Officials in Sport (OiS) – referees, umpires, judges and stewards are an often-studied profession for naturalistic researchers (Hancock & Ste-Marie, 2014; MacMahon & Plessner, 2008; Mallo, Frutos, Juárez, & Navarro, 2012; Mascarenhas, Collins, Mortimer, & Morris, 2005; McLennan & Omodei, 1996; Rix-Lievre, Recope, Boyer, & Grimonprez, 2013). Studying OiS performance allows researchers to understand phenomena which happen in the real world (G. Klein, 1993, 1998, 2008; Orasanu & Connolly, 1993). Further, naturalistic research in sport, including the study of OiS, emphasizes the way individuals, teams and systems conduct tasks, meet goals and make decisions (Kermarrec & Bossard, 2014; G. Klein, 1998; Macquet, 2009; Macquet & Fleurance, 2007; Orasanu & Connolly, 1993). Despite the previous studies mentioned, little research has explicitly examined decision making in OiS teams (Neville & Salmon, Under Review).

The Recognition Primed Decision (RPD) model, with its genesis in the study of how a fire ground commander (fire fighter) makes a rapid decision, establishes that decisions are made through priming (Gary Klein, Calderwood, & Clinton-Cirocco, 2010). In the RPD model an individual or team will have sufficient knowledge or expertise in the task to quickly match or compare different options. The model describes three approaches or variations for naturalistic decision making: Variation one – Simple Match (V1) states that experts will rely on experience and intuition to quickly match and select the most appropriate option. Variation two - Simulation/Diagnose(V2) occurs when no immediate option exists and the decision maker is required to spend time simulating or evaluating different options. Finally, variation three - Evaluate (V3) is used when modifications are required to possible decision actions in order for them to work (G. Klein,1993).

Despite the lack of research focussing on decision making in OiS, studying the RPD model in sport has provided insights into the performance of players in Basketball, Football, Handball and Volleyball. Results have identified that different sports exhibit a similar breakdown between the three variations (for a summary see Kermarrec and Bossard (2014)). Typically, participants have been found to exhibit V1, or 'simple match' strategies between 80-85% of the time (Kermarrec & Bossard, 2014). In a recent study, however, results identified that defensive soccer players, who are required to be more reactive than proactive when making decisions, have a V1/V2/V3 of 60%/24%/16% spilt with less of a reliance on V1 (60%) compared to players in other sports (Kermarrec & Bossard, 2014).

When investigating OiS performance, naturalistic research has examined the factors impacting decision making. For example, studies have examined the impact of factors such as emotion (Rix-Lievre et al, 2013), training (Mascarenhas et al, 2005), and positioning (Mallo et al, 2012) on decision making. Recent studies have also examined the role of team cognition in decision making (e.g. Boyer et al, 2013). The extent to which OiS follow similar decision making strategies to players is not clear. For example, when players on a field have more time to consider their options the frequency of simulation (V2) increases. (Kermarrec & Bossard, 2014). Are officials able to simulate or diagnose a situation and make a decision or does their role as instant arbiters of rule infringements lead to a dominance of V1?

OiS are constantly making adjudications of the game, performing a number of tasks, decisions and non-decisions in a rapid nature. Like the players in the game, OiS are performing in a dynamic environment under multiple stresses. While Cardin, Bossard, and Buche (2013) state that in sport the "quality of decision is ... seen as the ability of an athlete to act at any moment of the game quickly and efficiently;" for OiS performance and effectiveness is influenced as much by time as it is by accuracy. This subtle difference occurs because the dynamics of the environment for the umpire are manifestly different. The players acting as "protagonists" (Cardin et al., 2013) have their behaviour and decision making influenced by many pressures, including time, game situation and desire to beat the opposing team; Umpires, as the supporting cast, need only to adjudicate accurately without unduly delaying the protagonists' ability to continue the game.

The aim of this paper is to present the findings from an exploratory study of decision making in Australian Rules Football (AFL) officials. The aim of the study was to examine verbal in-game communication of umpires and determine if the communication represents how a decision is made with respect to the RPD model. Finally, the study aimed to identify if AFL umpires decision making follows a similar pattern to players or if, due to their role in the game as the support cast, their decision making is different.

Officiating Australian Rules Football

AFL is a fast paced ball sport where two teams of 18 players compete to score points (through goals and 'behinds') on an oval shaped field (illustrated in Figure 1) over four quarters of twenty minutes playing time. Players are required to kick an oval shaped ball through the Goals to score six points; if they miss to either side of the Goals (known as Behinds) or if the ball is not kicked through the Goals or Behinds the attacking team scores one point. There is no off-side and players hold notional positions as forwards (six players), midfielders (six players) and defenders (six players) (Australian Football League, 2014). The field is divided, through ground markings, into 'zones' known as attacking 50m arc, centre square and defending 50m arc. Players advance the ball towards the opposition's goal via kick or hand passes. A 'mark' occurs when a receiving player catches a kicked pass of over 15 metres in length before the ball hits the ground. While there are numerous infringement rules, at the basic level the laws of the game ensure that the players head, shoulders, back and legs are protected; while attacking players are limited in how they are required to dispose of the ball, with illegal disposals also adjudicated.

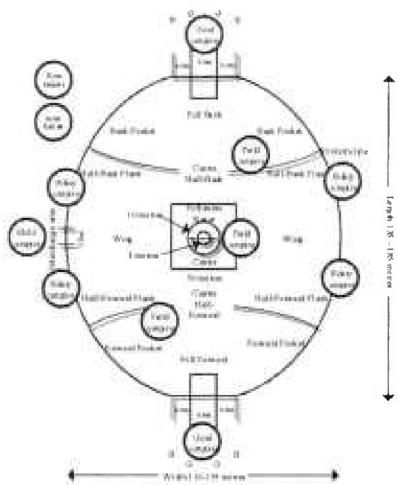


Figure 1 - AFL field dimensions with notional playing and umpiring positions (Australian Football League, 2014)

There are nine on-field umpires in AFL: three Field Umpires, one in each zone. four Boundary Umpires and two Goal Umpires, one at each end. The on-filed umpiring team is supported at ground level by a reserve Field Umpire (known as the Emergency Umpire) and reserve Goal Umpire as well as, in the stands, official timekeepers and a score review umpire who has access to instant replays.

The Filed Umpires adjudicate all marks and rule infringements. An indication of a mark by an umpire allows a player to stop the flow of play and take an unpressured kick. An umpire is required to decide if the ball has been touched and if it has travelled the required distance for a mark to be paid. When an umpire decides a rule infringements has occurred a Free Kick is awarded to the offended team. If a contest is adjudged fair and within the rules, the umpire will call 'play on.'

The Boundary Umpires adjudicate when a ball leaves the field of play. If the ball leaves the playing surface 'on the full' from a players foot then the opposition is awarded a Free Kick, termed Out of Bounds on the Full (OOBF), in (nearly) all other cases the boundary umpire will return the ball to play through a Boundary Throw In (BTI).
The Goal Umpires adjudicate when a ball goes through the Goals or Behinds. The Goal Umpires also keep the official score of the game and provide scoring signals to the players, umpires and spectators. Goal Umpires are assisted by Boundary Umpires, Field Umpires and a video review system (termed Score Review) for scoring decisions if ambiguity exists.

Field and Boundary Umpires communication to the players using a whistle and short verbal communications. Typically, a single whistle is used to stop the flow of the game, while a double whistle is used to get a players attention when the game has been stopped. Through the course of a game the three Field Umpires, two Goal Umpires, the Emergency Umpire and timekeepers use a radio communication system to provide open, real-time continuous communication. The study, while interested in all member of the AFL umpiring team, focused its research on the three Field umpires.

METHOD

Data Collection

The Australian Football League Umpiring Department provided the authors with audio/visual recordings of three AFL games from the 2014 AFL Premiership Season. The vision of the game was the same as the host television broadcasters; the audio tracks, however, had the television commentary removed providing an uninterrupted stream of the umpires' communications which were recorded via the standard match communications system currently worn by AFL umpires.

Participants

For the purpose of the study, the subjects were eight male AFL Field Umpires. The combined experience of the umpires was 928 (μ - 116; σ -77.8) AFL games at the beginning of the 2014 AFL Premiership Season. Due to manner in which the data was obtained it was not possible to gather information regarding the age of the participating umpires.

Materials

The communication made by the umpires was captured by the radio communications equipment and recorded in synchronisation with the live television broadcast onto a DVD. The VLC media player and Microsoft Excel was used to conduct the transcription and data analysis.

Data Analysis

Each game was transcribed verbatim from the recoded footage. One analyst then reviewed the footage and transcripts to identify instances of communication during the games. Each umpire communication was transcribed as a 'communication instance', defined as a word, phrase or use of the whistle made by an umpire or any other communication picked up by the umpires' microphone. For each communication instance the game, quarter and time stamp was recorded as well as the area on the playing field where the communication occurred.

Each communication instance was coded to identify sequences of communications which represented decision moments in the game. A decision moment was defined as a moment when a field umpire had to decide to intervene on the game; to inject a decision which would alter the regular flow of the game.

The decision moments, as a verbal record of decision made by the umpiring team, were then coded by one analyst as one of the three variations in the RPD model as presented by (G. Klein, 1993, 1998). Coding occurred through assessing the set of communication instances contained within each decision moment. For reliability purposes a second analyst also coded the decision moments as one of the three RPD variations. Comparison of both analysts coding revealed an agreement of 94%. For those decision moments on which the analysts did not agree consensus was achieved through further discussion.

RESULTS

Frequency of Communication Instances

Table 2 shows frequency counts of the 15 most recorded communication instances across the three games analysed. In total 6025 communication instances were identified of which there 960 unique and 64 were repeated 10 or more times. The use of the whistle is the most frequent communication instance followed by the use of the verbal play on instruction and inter-umpiring team control instructions such as 'you' and 'me'. The control instructions are used to indicate, within the team, which umpire is responsible for making decisions at any one moment.

Decision Moments

Table 3 presents a breakdown of the decision moments identified in the transcribed data. The decision moments are separated based on the type of event in the game that they are related to - marks, Free Kicks and stop ball situations (Ball Ups, ball out of play moments and scoring resets).

Table 2. Frequency of the most used communication instances per quarter (Q1...Q4) and game (G1,G2,G3). Communications instances in italics represent umpire to umpire communications.

	Game 1					*Game 2*					*Game 3*					**Total**
Communication Instance	Q1	Q2	Q3	Q4	**Total**	Q1	Q2	Q3	Q4	**Total**	Q1	Q2	Q3	Q4	**Total**	
Whistle	76	98	81	83	**338**	101	89	90	84	**364**	93	93	99	84	**369**	**1071**
'Play on'	85	86	71	94	**336**	91	81	78	57	**307**	97	77	83	89	**346**	**989**
'You'	39	31	32	32	**134**	40	33	30	27	**130**	37	28	35	31	**131**	**395**
'Me'	35	37	44	36	**152**	14	19	13	28	**74**	28	23	32	32	**115**	**341**
'Thank you'	12	6	6	4	**28**	7	14	15	12	**48**	24	16	21	33	**94**	**170**
'Yep'	16	12	16	12	**56**	14	13	16	13	**56**	12	16	12	15	**55**	**167**
'All clear'	17	11	14	9	**51**	7	13	17	12	**49**	11	15	13	11	**50**	**150**
Double whistle	9	7	8	11	**35**	12	12	8	12	**44**	6	6	13	9	**34**	**113**
'Thanks [player]'	5	9	7	2	**23**	2	9	9	14	**34**	7	8	5	10	**30**	**87**
'[player]'	4	1	4	-	**9**	3	6	11	14	**34**	6	14	10	4	**34**	**77**

Goal restart	10	6	9	6	**31**	4	3	11	5	**23**	6	5	6	6	**23**	**77**
'Move it on'	2	4	6	7	**19**	10	8	8	5	**31**	4	4	12	6	**26**	**76**
'Good [umpire]'	8	12	10	10	**40**	6	1	5	1	**13**	3	4	3	4	**14**	**67**
'Backing back'	12	3	9	6	**30**	6	4	11	10	**31**	-	-	2	2	**4**	**65**
'Mark's here'	7	3	4	1	**15**	1	2	4	5	**12**	5	9	3	5	**22**	**49**

Table 3. Decision moments instances per quarter (Q1...Q4) and game (G1,G2,G3)

	Game 1					*Game 2*					*Game 3*					**Total**
Decision Moment	Q1	Q2	Q3	Q4	**Total**	Q1	Q2	Q3	Q4	**Total**	Q1	Q2	Q3	Q4	**Total**	
Mark	35	40	36	44	**155**	47	49	35	32	**163**	50	43	38	49	**180**	**498**
BTI	8	15	10	9	**42**	15	8	7	12	**42**	12	15	12	7	**46**	**130**
Free	11	10	10	8	**39**	9	6	11	9	**35**	7	6	12	7	**32**	**106**
Kick In	6	5	4	4	**19**	3	8	5	7	**23**	6	10	6	6	**28**	**70**
Ball Up	4	2	7	5	**18**	3	5	8	9	**25**	4	3	7	6	**20**	**63**
OOBF	1	2	-	1	**4**	2	1	1	1	**5**	-	2	-	3	**5**	**14**
Recall	-	-	-	-	-	-	2	1	-	**3**	-	-	-	1	**1**	**4**
SR	-	-	-	-	-	-	-	-	-	-	-	-	1	-	**1**	**1**
Other	-	-	-	-	-	-	-	1	-	**1**	-	-	-	-	-	**1**
Total	**65**	**74**	**67**	**71**	**277**	**79**	**79**	**69**	**70**	**297**	**79**	**79**	**76**	**79**	**313**	**887**

As shown in Table 2, just over 56% of the decision moments were mark decisions whereby the umpire determines whether a mark has been made. The next most frequent decision moment across the three games related to Boundary Throw In decisions, followed by Free Kicks, Kick Ins, and Ball Up decisions.

RPD Variation

Tables 3 and 4 present the number of decision moments associated with each variation in the RPD – V1 - simple match, V2 - simulate/diagnose and V3 - evaluate. Table 4 presents the variation breakdown for each game per quarter, while Table 3 shows that 78% of the decisions occurring across the three games were characteristic of V1, with just over 18% representing V2 and 3.5% V3.

Table 5 presents the breakdown against the different decision moments.

Table 4. Breakdown of RPD Variation by quarter

	Variation 1 (Simple Match)		**Variation 2 (Simulate/diagnose)**		**Variation 3 (Evaluate)**		
	Raw	%	Raw	%	Raw	%	**Total**
Game 1							
Q1	49	75.4%	14	21.5%	2	3.1%	**65**
Q2	60	81.2%	12	16.2%	2	2.7%	**74**
Q3	50	74.6%	17	25.4%	-	-	**67**
Q4	56	80.3%	10	14.1%	4	5.6%	**71**
Game Total	**216**	**78.0%**	**53**	**19.6%**	**8**	**2.9%**	**277**
Game 2							
Q1	64	81.%	11	13.9%	4	5.1%	**79**
Q2	65	82.3%	13	16.5%	1	1.3%	**79**
Q3	44	64.7%	17	25.0%	7	10.3%	**68**
Q4	51	72.9%	17	24.3%	2	2.9%	**70**
Game Total	**224**	**75.7%**	**58**	**19.6%**	**14**	**4.7%**	**296**
Game 3							
Q1	68	86.1%	11	13.9%	-	-	**79**
Q2	67	84.8%	11	13.9%	1	1.3%	**79**
Q3	56	73.7%	13	17.1%	7	9.2%	**76**
Q4	62	78.5%	16	20.3%	1	1.3%	**79**
Game Total	**253**	**80.8%**	**51**	**16.3%**	**9**	**2.9%**	**313**
Total	**693**	**78.2%**	**162**	**18.3%**	**31**	**3.5%**	**886**[10]

Table 3 shows that 78% of the decisions occurring across the three games were characteristic of V1, with just over 18% representing V2 and 3.5% V3.

[10] The decision moment coded as 'other,' a scuffle between players, did not require an intervention by the umpires and was not included in the RPD analysis.

Table 5. Breakdown of RPD variation by decision moment

	Variation 1 (Simple match)		Variation 2 (Simulate/diagnose)		Variation 3 (Evaluate)		
	Raw	%	Raw	%	Raw	%	**Total**
Mark	495	99.4%	-	-	3	0.6%	498
BTI	129	99.2%	-	-	1	0.8%	130
Free Kick	-	-	83	78.3%	23	21.7%	106
Kick In	69	98.6%	-	-	1	1.4%	70
Ball Up	-	-	61	96.8%	2	3.2%	63
OOBF	-	-	14	100%	-	-	14
Recall	-	-	4	100%	-	-	4
Score Review	-	-	-		1	100%	1
Total	**693**	**78.2%**	**162**	**18.3%**	**31**	**3.5%**	**886**

Table 4 shows that the majority of VI decisions were marks and game resets ('Boundary Throw In' and point 'Kick In'). V2 decisions were split between game resets ('Ball Up' and 'Out of Bounds on the Full') and Free Kicks. Free Kicks were also the most dominant decision moment in V3.

DISCUSSION

This study is the first in a sequence examining the nature of AFL umpire decision making during three elite level AFL games. The findings provide some interesting points around the characteristics of AFL umpire decision making. In the following discussion these are discussed through a RPD lens.

Verbal articulation of AFL Umpires

The use of the whistle was the most prominent verbal articulation in the data; providing clear moments where the umpires intervenes and verbalizes their decision making. The whistle is used to indicate an intervention in the game – either a mark or Free Kick. It is also used as a form of call and response between Field and Boundary umpires to first indicate and then acknowledge that the ball has gone out of play. Finally, it is used to encourage a player to move the ball on and restart play (Double whistle) after an intervention moment.

The verbal articulation to play on was the next most common communications instance. Play on was used an indicator to the players that they play is live, that the game can continue and that no rule breach has occurred. That is, when an umpire encountered situations (or contests) where two or more players were legally competing for the ball the umpire called play on to inform them that no intervention would occur. Using play on in this context is considered a non-decision, where a decision to not intervene is made and then verbalized to the players. Play on is also used to inform players that, after an intervention moment, the game is live again and that a contest between competing players is permitted. In the data the use of play on in these contexts was not universal as not every non-decision or non-intervention moment required the umpire to call play on.

The prevalent use of play on shows that when an umpire is primed by a contest between two players the default (or simple match) of the umpire is to let play continue and call play on. It is only after a contest does not meet the play on criteria, a rule breach for example, that an umpire intervenes through the use of the whistle. Although it was not possible to identify all instances of non-intervention, the count of play on (989) noticeably outnumbers any single type of decision moment, suggesting that when an umpire is primed by a contest between two players their simple match is to call play on.

The RPD model applied to AFL Umpires

Analysis of the decision moments indicated that, in the three games analysed, the umpires followed a 78.2/18.3/3.5% split between the three variations in the RPD model. The split suggests that the majority of AFL umpire decision making comprises VI or simple match decisions. For VI decisions (78.2%) analysis showed that the majority of the mark, boundary throw in and kick in decision moments conformed to the simple match criteria. With a mark, an umpire is primed by the kicking of the ball by one player and pays a mark if the ball is caught without anything complicating the situation. Similar priming criteria exits when the ball crosses the boundary or is kicked in following a behind being scored.

Decision moments using V2 (18.3%) included three distinct contexts – Free Kick, Ball Up and OOBF. For a Free Kick, an umpire has to adjudicate that a rule breach has occurred and that an intervention is required. While it may be possible that a contest between two players prime an umpire towards a rule breach the frequent use of play on reveals that an umpire's default strategy is to let the play continue; implying a consideration of another option, such as a rule breach.

More interestingly, the ball up decision moment, where an umpire decides that the play has stopped and needs to be reset (similar to jump ball in basketball or a drop ball in soccer) reveals a verbal simulation of options before a decision action. Within a ball up decision moment an umpire can be heard calling play on for several contests before the whistle is sounded for a ball up. For example one decision moment contain the following instances "Play on, play on, play on, [whistle], my ball, I'll have it." Each use of play on indicates that the umpire has considered the contest to be fair. After the third play on call the umpire decided the ball had stopped moving and the play needed to be reset. In this respect the umpire has taken time to consider different options to intervene before deciding to act. The OOBF decision moment occurs when the ball is kicked out of bounds without being touched or hitting the ground, a trigger for a Free Kick to the opposing team. OOBF instances, as seen in Table 2, occur infrequently in a game; and, similar to the Free Kick and ball up decision moments, the umpire first considers how the ball has left the playing field before deciding on OOBF.

Finally, the small number of V3 (3.5%) decision moments occurred when the verbal communication provided by the umpiring team indicated changes to the original course of action. The intervention of another umpire suggests that the umpiring team had to engage in further evaluation of the situation following an original V1 or V2 decision. In the instance of a mark decision moment, for example, the course of action for a regular mark changed when a defending play interfered with the attacking player after the mark was taken; resulting in a metreage penalty being applied to the defending team. For Free Kicks, instances included players electing to take advantage or a non-controlling umpire paying a Free Kick in the zone of the controlling umpire. The V3 Free Kick decision moments resulted in umpires having to communicate an alternate course of action to that which happen for regular Free Kicks, indicating modifications to the decision actions of an umpire.

It is notable that the findings from the present study are similar to those that have examined decision making in other sports. The finding that 78% of decision moments were characteristic of the V1 - simple match variation of RPD, is similar to the findings from studies in Volleyball (Macquet, 2009) and Ice Hockey (Bossard, De Keukelaere, Cormier, Pasco, & Kermarrec, 2010; Mulligan, McCracken, & Hodges, 2012); however, it is notable that this is the first study to examine OiS as opposed to players. The high percentage of V1 decisions indicates that as instant arbiters of rule infringements, AFL umpires decision making is dominated by simple matches. When comparing the split between V2 and V3 it was identified that AFL Umpires favour V2 (18.3%). The relatively low proportion ascribed to V3 (3.5%) may be accounted by the fact that AFL umpires, as support actors in the game, have less capacity, either through time or other stressors, to modify existing decision actions.

Team decision making

An interesting phenomena observed through the analysis of the data was the degree to which members of the umpiring team were able to coordinate decision making through the use of the communications equipment. In the decision moments which conformed to V3 the data identified multiple umpires providing instructions to the controlling (or deciding) umpire. It is possible that such communication facilitated more accurate decision making due to the pooling of the experience and perspectives of the other team members. Additionally, the data has identified that the communications technology has allowed the entire team to know what is going on where 'the play' was occurring. In relation to the RPD model, this finding demonstrates the important role that teamwork, communications and communications technology have to play in V3 decision making. V1 and V2; however, can proceed as an individual function.

Further Research

Whilst further examination of OiS decision making across sports is recommended, a limitation of the present study is that it examined decision making through umpire communications and game vision only, without obtaining the perspective of the decision maker. The limitation can be overcome through conducting self-confrontation interviews using techniques such as critical decision method or verbal protocols with the participants.

Self-confrontation interviews will allow researchers to understand the impact the communications technology has on the decision making processes. Further research is also required into the degree to which AFL umpires conduct a single decision action or implement multiple successive decision actions using some form of anticipatory thinking

(G Klein, Snowden, & Pin, 2011). Does a single decision moment consist of an umpire stepping between the different variations of the RPD model? Self-confrontation interviews currently appear to be the only way to test such a hypothesis.

As the audio feed in the original data has been synchronized with vision of the game, the ability to time stamp each communication instance and decision moment provides a rich data set to enable quantification of the rapid decision making of the umpires. While it is commonly accepted that the RPD model describes how a rapid decision is made (G. Klein, 1993), the combination of verbalized in-the-moment decision making and time stamp communication instances allows for an understanding of how quick a rapid decision is made in the context of AFL and sport.

CONCLUSIONS

It is concluded that umpiring in AFL involves all three variations of RPD; however, the majority of decisions reflect V1 – simple match. In addition, despite significant differences across sports, there appear to be similarities between the breakdown of RPD variations adopted by officials in AFL and players in other sports such as volleyball and ice hockey. Suggesting that, despite their role as supporting actors in sport, OiS make use similar decision making strategies.

In relation to AFL, it is concluded that the umpires verbalize a number of non-decision 'play on' calls through the course of a game, implying that umpires are primed to not make a rule adjudication (and award a Free Kick). AFL umpires also demonstrated an ability to use communication technology as a means to evaluate more complex decision moments.

Using real time in-game communication technology has enabled researchers to understand the naturalistic decision making without any direct intervention in the tasks being undertaken. Further, due to the training and tasks requirements of AFL Umpires, the verbalized data has provided a unique way to capture 'in-the-moment' decision making.

Finally, this study demonstrates how research into the naturalistic decision making of OiS provides a low risk non-invasive domain to test, explore and extend decision making and team work models, which in turn can be translated to other safety critical areas.

ACKNOWLEDGEMENTS

This study was conducted with the support of the Australian Football League Umpiring Department. Timothy Neville receives funding from the Defence Science and Technology Organisation and the University of the Sunshine Coast. Paul Salmon receives funding from the Australian Research Council. The authors wish to thank Dr Alexander Kalloniatis for his support in reviewing the manuscript.

REFERENCES

Australian Football League. (2014). Laws of Australian Football 2014. In A. F. League (Ed.). Melbourne, Victoria: Australian Football League.

Bossard, C., De Keukelaere, C., Cormier, J., Pasco, D., & Kermarrec, G. (2010). L'activité décisionnelle en phase de contre-attaque en Hockey sur glace [Decisional activity during counterattacks in ice hockey]. *@ctivités, 7*(1), 41-60.

Cardin, Y., Bossard, C., & Buche, C. (2013). *Investigate naturalistic decision-making of a workgroup in dynamic situation. From the modelling to the design of a training virtual environment.* Paper presented at the NDM11, the 11th International Conference on Naturalistic Decision Making, Marseille, France,.

Hancock, D. J., & Ste-Marie, D. M. (2014). Describing Strategies Used by Elite, Intermediate, and Novice Ice Hockey Referees. *Research Quarterly for Exercise and Sport, 85*(3), 351-364. doi: 10.1080/02701367.2014.930090

Kermarrec, G., & Bossard, C. (2014). Defensive Soccer Players' Decision Making: A Naturalistic Study. *Journal of Cognitive Engineering and Decision Making.* doi: 10.1177/1555343414527968

Klein, G. (1993). A recognition-primed decision (RPD) model of rapid decision making. *Decision making in action: Models and methods, 5*(4), 138-147.

Klein, G. (1998). *Sources of power : how people make decisions / Gary Klein*: Cambridge, Mass. : MIT Press, c1998.

Klein, G. (2008). Naturalistic decision making. *Human Factors, 50*(3), 456-460. doi: 10.1518/001872008X288385

Klein, G., Calderwood, R., & Clinton-Cirocco, A. (2010). Rapid Decision Making on the Fire Ground: The Original Study Plus a Postscript. *Journal of Cognitive Engineering and Decision Making, 4*(3), 186-209. doi: 10.1518/155534310x12844000801203

Klein, G., Snowden, D., & Pin, C. L. (2011). Anticipatory thinking. *KL Mosier, & UM Fischer, Informed by Knowledge*, 235-246.

MacMahon, C., & Plessner, H. (2008). The sport official in research and practice. *Developing Sport Expertise*, 172-192.

Macquet, A. C. (2009). Recognition within the decision-making process: A case study of expert volleyball players. *Journal of Applied Sport Psychology, 21*(1), 64-79. doi: 10.1080/10413200802575759

Macquet, A. C., & Fleurance, P. (2007). Naturalistic decision-making in expert badminton players. *Ergonomics, 50*(9), 1433-1450. doi: 10.1080/00140130701393452

Mallo, J., Frutos, P. G., Juárez, D., & Navarro, E. (2012). Effect of positioning on the accuracy of decision making of association football top-class referees and assistant referees during competitive matches. *Journal of Sports Sciences, 30*(13), 1437-1445. doi: 10.1080/02640414.2012.711485

Mascarenhas, D., Collins, D., Mortimer, P., & Morris, R. (2005). A naturalistic approach to training accurate and coherent decision making in rugby union referees. *The Sport Psychologist, 19*(2), 131-147.

McLennan, J., & Omodei, M. M. (1996). The role of prepriming in recognition-primed decisionmaking. *Perceptual and Motor Skills, 84*(3 PART 2), 1059-1069.

Mulligan, D., McCracken, J., & Hodges, N. J. (2012). Situational familiarity and its relation to decision quality in ice-hockey. *International Journal of Sport and Exercise Psychology, 10*(3), 198-210.

Neville, T. J., & Salmon, P. M. (Under Review). Never Blame the Umpire - A Review of Situation Awareness Models and Methods for Examining the Performance of Officials in Sport *Ergonomics*.

Orasanu, J., & Connolly, T. (1993). The reinvention of decision making. In G. A. Klein, J. Orasanu, R. Calderwood & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 3-20). Westport, CT, US: Ablex Publishing.

Rix-Lievre, G., Recope, M., Boyer, S., & Grimonprez, M. (2013). *Naturalistic decision making and emotion in refereeing: affect at the heart of judgment.* Paper presented at the NDM11, the 11th International Conference on Naturalistic Decision Making, Marseille, France.

Crowdsourcing Mental Models using DESIM (Descriptive to Executable Simulation Modeling)

Mark S. PFAFF, Jill L. DRURY, and Gary L. KLEIN
The MITRE Corporation

ABSTRACT
This paper describes the DESIM (Descriptive to Executable Simulation Modeling) process for transforming causal descriptive models into computer simulation models based on information obtained from crowdsourcing. Feedback obtained from crowdsourcing is used to quantify the strength of causal relationships between variables in descriptive models to provide an unbiased distribution of estimated weights for each causal relationship and thereby enable mathematical processing of the descriptive models on a computer. The approach employs fuzzy cognitive modeling methods to elicit and structure the models and the analytic hierarchy process to compute the distribution of weights between variables. The output of this process produces a decision space, which is visualized with a novel decision space visualization tool. An experimental application of this process is presented and discussed, with implications for future research.

KEYWORDS
Mental models; General and miscellaneous; Judgment and Decision Making; Planning and Prediction; Expertise

INTRODUCTION

There is increasing research on better ways to support decision makers when they need to choose among options in complex situations. Because of the deep uncertainty surrounding such situations (Walker, Lempert, & Kwakkel, 2012), questions arise about planning for the range of conditions under which reasonable operations would be possible, or determining what operations may be managed under current conditions (Caldwell, 2014). In some situations, decision makers match salient cues presented by the external environment to a mental template built from previous experiences: part of their mental model (Craik, 1943). They then envision at least one possible course of action and mentally simulate the results of applying that action to determine whether that option is acceptable (Klein, 1998). This process is more difficult to use in complex situations, or by novice decision makers who have not yet acquired sufficient experience to map the current situation to mental templates showing successful resolutions of problems faced in the past. Addressing this gap are *decision spaces*, defined as the range of options, the underlying interconnected factors that influence their relative desirability, and the landscape of plausible futures that could accompany any given course of action (Pfaff et al., 2012). For example, alternative courses of action can be evaluated using a computer simulation, which can provide decision makers with a graphical depiction of their decision space and thereby option awareness during real-world emergency situations, such as a natural disaster. The idea is that visually depicting decision spaces can be like providing night vision goggles for the mind, offloading the mental simulation process onto the computer, which displays the results of many posssible options using an intuitive decision space visualization (DSV) that otherwise cannot be seen unaided.

In laboratory experiments, DSVs have enabled decision makers to make choices faster, more accurately, and with more confidence than without the DSV (Pfaff et al., 2012). The process of creating DSVs relies upon exploratory modeling (Bankes, 1993; Chandrasekaran & Goldman, 2007; Chandrasekaran, 2005). In exploratory modeling, analysts construct a set of plausible assumptions about the environment in which a decision will be made, run a simulation model that includes a parameterization of those assumptions, and score the outcomes for each decision option according to one or more evaluative criteria. The analyst varies each of the parameters representing the assumptions to account for uncertainty, and runs the model repeatedly to obtain a range of outcomes for each decision option. DSVs consist of a frequency-based depiction of the range of outcomes for each option. Thus, the process of constructing a DSV requires executing a model that pertains to the domain and situation encountered by the decision maker. While there is a rich history of research into modeling and simulation, the fact remains that developing validated models can be costly, time-consuming, and error-prone. The need for models has become the stumbling block for creating DSVs for broad classes of decision making situations.

There is a need for a more streamlined way to develop models. Building from decision makers' mental simulation abilities, new research in crowdsourcing points to the promise of combining the mental models from multiple decision makers to paint a more complete and unbiased picture of a situation than any individual might be able to achieve in isolation. The result is a new process, DESIM (Descriptive to Executable Simulation Modeling), that can create a computational model of a situation in hours or days instead of weeks or months.

DESIM consists of the following stages:

- Create one or more validated descriptive causal models
- Deconstruct the model into pairwise comparisons
- Crowdsource the comparisons and compute relationship strengths
- Apply the computational model to create DSVs

The DESIM process transforms descriptive causal models into computer simulation models based on information obtained from crowdsourcing. A computer user interface for crowdsourcing, combined with computational algorithms, produces quantitative values for the strengths of causal relationships between variables in the descriptive models, resulting in unbiased distributions of estimated values for each relationship, and enabling the models to be computationally processed. This system generates improved outcome spaces, which refer to one or more possibilities regarding the relationships among options, actions, or variables that can be used to analyze the subject of a computer model. For example, a decision space can be an outcome space used by decision makers to determine how to respond to a complex situation based on the relationships between options and their plausible effects that can be forecasted by a computer simulation from facts about the situation. A key distinguishing feature of this system from other decision support tools is that it presents to the user an interactive and dynamic frequency distribution of possible outcomes (e.g. box-plot or histogram) rather than a single static probability, which conceals important knowledge about the range and distribution of possible outcomes. For example, perceiving a distribution with a long tail or a bi-modal distribution may lead to a significantly different decision (or to further exploration of the data) than simply knowing the mean probability of success. Moreover, in this format, further exploration of the data can yield deeper awareness of what factors lead to better vs. worse outcomes (Drury, Klein, Musman, Liu, & Pfaff, 2012). The rest of this paper describes how the DESIM process works, prefaced by related work and ending with a brief example of using DESIM.

RELATED WORK

Conventional computer simulation systems include models that are designed based on expert knowledge for use to simulate different situations that may occur in the real-world. These computer simulations are assumed to be reliable because they are created using expert knowledge. Unfortunately, each expert has certain behavioral patterns, preferences, and characteristics that may bias the programming of models. For example, different experts may agree to include certain variables in a particular computer model but disagree about the significance of each variable. Thus, conventional computer simulations created using expert models may be biased and unreliable, and the process of translating expert knowledge into computer simulation models can be slow and error prone (Bankes, 1993). Because this approach most often requires computer programmers to do this translation, the process is slow and tedious since the translation must be carefully and constantly validated by the experts to eliminate translation errors.

Alternatively, a domain expert's (such as an analyst or forecaster) descriptive causal model can be elicited for a focal question such as "Will Iran invade the Strait of Hormuz?" and then represented in a digital data structure that is interpretable by computer programs and, moreover, can be displayed in a graphical presentation that is easy for the expert to validate. This is the approach taken with DESIM. Most often, multiple domain experts are interviewed to understand different and potentially conflicting perspectives on the problem. Experts identify model components, links between components, and the dynamic and functional relationships among the components. The advantages of this process include the ability to capture perceptions that are difficult to quantify and the participatory form makes data collection more approachable and engaging for domain experts not familiar with modeling processes (Özesmi & Özesmi, 2004). An alternative approach is to separate the work of developing the formal model from the interview process, such that the model itself is designed by modeling experts, based on the knowledge collected from interviews with domain experts. The resulting model is then validated by displaying a graphical representation to the domain experts who check it for completeness and accuracy (Sieck, Rasmussen, & Smart, 2010).

There are multiple tools for computationally representing mental models. The model in Figure 1, below, was designed using CMapTools (Cañas et al., 2004), which provides a graphical interface for constructing and editing cognitive models and provides machine-readable output for use by other computational tools. The edge labels are an open text field which here is used to signify positive or negative associations between nodes. A similar product is MentalModeler (Gray, Gray, Cox, & Henly-Shepard, 2013), designed to support the fuzzy cognitive modeling process from model elicitation and graphical representation, variable edge weight selection, to exploratory simulation (it also can export the model in a machine-readable form).

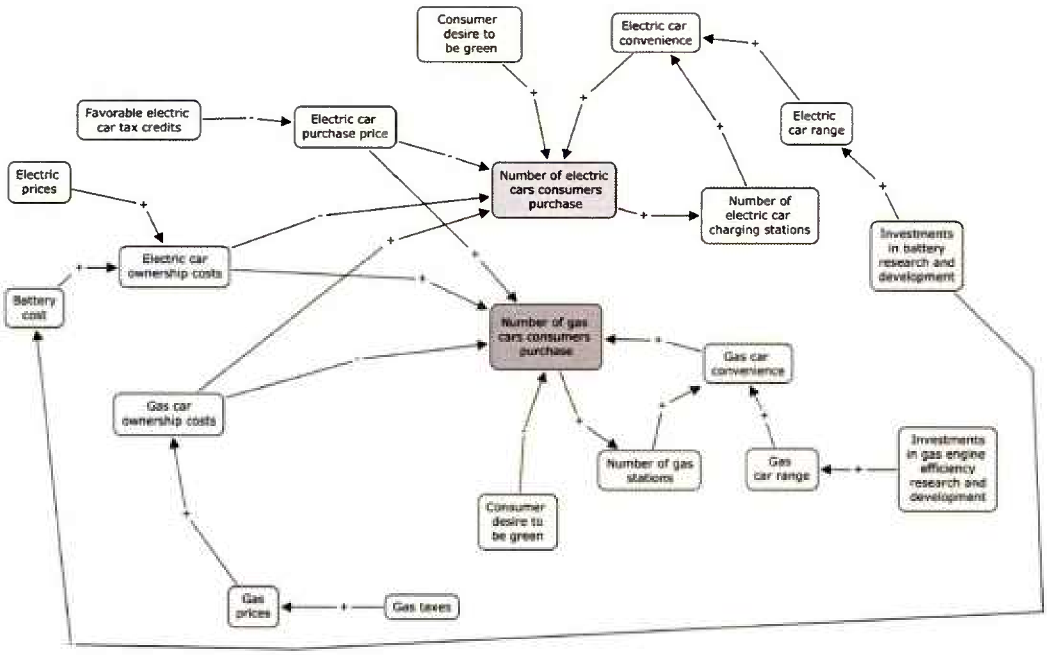


Figure 1. Example cognitive model describing reasons to purchase a gas or electric vehicle

DESIM also uses crowdsourcing (or crowd estimating) to quantify the relationships in the descriptive model Crowdsourcing is a process of obtaining services, ideas, or content by soliciting contributions, especially over the Internet, from a large group of people referred to as a crowd (Howe, 2006). This process typically involves a division of labor for tedious tasks split among members of the crowd. For example, crowdsourcing can be used to solicit predictions for a political campaign, or to search for answers, solutions, or a missing person (Surowiecki, 2005). In other words, crowdsourcing combines the incremental efforts of numerous contributors to achieve a greater result in a relatively short period of time. Lin et al. (2012) used crowdsourcing to understand mental models of privacy in mobile applications, but did not create a model explicitly.

METHODS

This section details the methods for eliciting the canonical cognitive model of the problem, representing the model interactively, eliciting values for the model from the crowd, and analyzing and visualizing the resulting data. In the first stage, analysts develop a focal question of interest and interview one or a few experts on the subject area to elicit the experts' mental models of the factors they believe influence the outcomes of the focus question. The analysts develop one causal model per expert in the form of nodes and unweighted edges, validating each model with its expert. The causal mental models may be similar or may diverge. Analysts combine multiple models when they are mathematically equivalent, but it is acceptable to have more than one canonical model describing the experts' mental models.

The descriptive causal model can be represented on a computer as a graph of nodes connected by edges, as shown in Figure 1, for which the focal question asked of domain experts was "Will consumers buy more electric vehicles than gas vehicles in 2018?" A node in a descriptive model is a variable that represents a concept such as an action, option, or policy that has a range of values. Different scenarios using the model would be described with different sets of initial node values. An edge includes a weight that represents a causal association or relationship between two or more nodes. The sign of an edge weight denotes a direction of correlation between nodes, and the magnitude of an edge weight denotes the strength of the causal relationship between the

nodes. While a static value for an edge weight may be elicted from a single expert, a distribution of values for the edge weights can be determined through appropriately crowdsourcing to multiple experts. An algorithm can be used to determine how crowd sourced feedback defines the distribution of edge weights (described further below). Once the canonical map is elicited from one or more domain experts and represented structurally, the DESIM computer program processes it in parts in order to obtain edge weights. While experts are able to give the sign (+ or -) of a causal relationship, they are less able to give an accurate estimate of the magnitude (Osei-Bryson, 2004). Because subjective point estimates are unreliable, another method is necessary to produce accurate edge weights. This is achieved through a systematic set of pairwise comparisons of the connected node pairs in the model, where an expert rates the comparative strength of two relationships (e.g. whether relationship X 1 → X2 is stronger than X 3 → X4, and by how much). In this study, crowdsourcing of the pairwise comparisons was used via Amazon Mechanical Turk (AMT) because the topic was of general interest and required no special expertise. More specialized models would require targeted recruitment of contributors, for example, from a community of analysts or forecasters in a specific field.

A web-based interface was built to elicit the online pairwise comparisons. This tool takes as input the machine-readable model produced in the preceding steps and generates the set of pairwise comparisons which are presented in sequence to the users recruited via AMT. First, a single relationship X1 → X2 (in this example X1 has a positive relationship with X2) is graphically presented to the user with the question "Do you agree with this relationship?" The three choices are "Agree: An increase on the left causes an increase on the right", "Disagree: An increase on the left has no effect on the right", or "Disagree: An increase on the left causes a decrease on the right" (see Figure 2).

After the user has agreed with at least two relationships in the model, two relationships (A = X1 → X2 and B = X3 → X1) are presented with the question "Which relationship is stronger?" with the choices "A is stronger than B," "B is stronger than A," or "A is the same as B." (see Figure 3). If either of the first two choices are selected, the user is additionally asked "How much stronger?" and presented with a slider ranging from "A is much stronger than B" to "A is the same as B." After answering, the user then proceeds to the next comparison. When the user disagrees with a given relationship, it is given a weight of zero and eliminated from all future pairwise comparisons. Because different users may agree or disagree with different relationships, the resulting set of pairwise comparisons will vary in their degree of completeness.

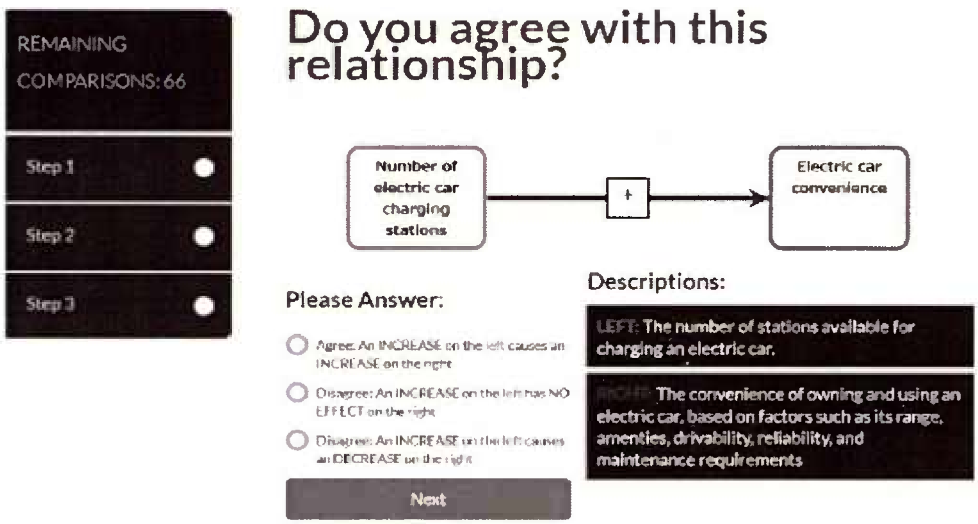


Figure 2. Step 1 of causal relationship pairwise comparison

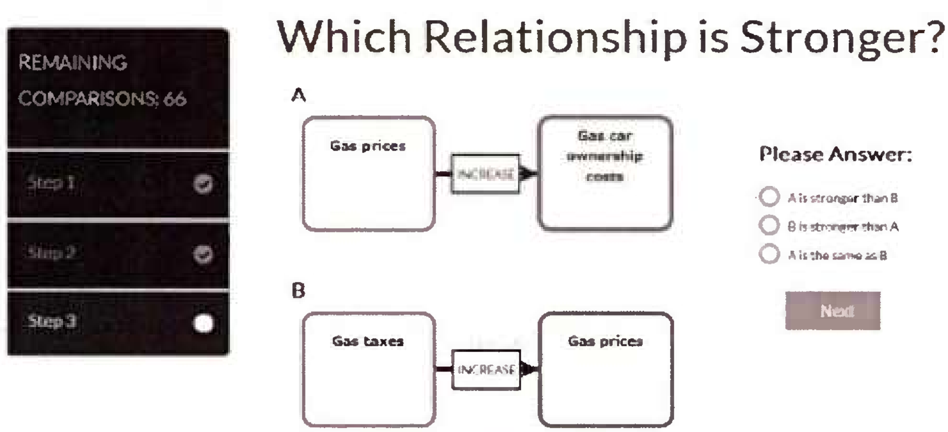


Figure 3. Step 2 of causal relationship pairwise comparison

DESIM analyzes the results of the pairwise comparisons using the Analytic Hierarchy Process (AHP; Saaty, 1990). From each respondent's set of pairwise comparisons, a set of edge weights is computed using a modified AHP technique to accommodate incomplete sets of pairwise comparisons (Harker, 1987). The sets of edge weights for all respondents (which are values between 0 and 1; the model defines the sign) are aggregated and used to populate the original model with distributions of edge weights for each relationship in the model.

Using these distributions of weights, multiple simulation model processing runs can be performed to generate one or more outcome spaces, using an iterative Fuzzy Cognitive Modelling (FCM) method (Kosko, 1986). Initial node values and edge weights can be varied for each model processing run to create an outcome node distribution. Variations may be generated in different ways. For example, in the example presented here, the model was executed once for each set of edge weights elicted from the respondents. Alternately, a Monte-Carlo method can be used to generate each variation by sampling from distributions of node values and edge weights. An analysis of the resulting outcome space provides a more comprehensive understanding than a single aggregated mean estimate about how various variables impact consumer interest in electric cars.

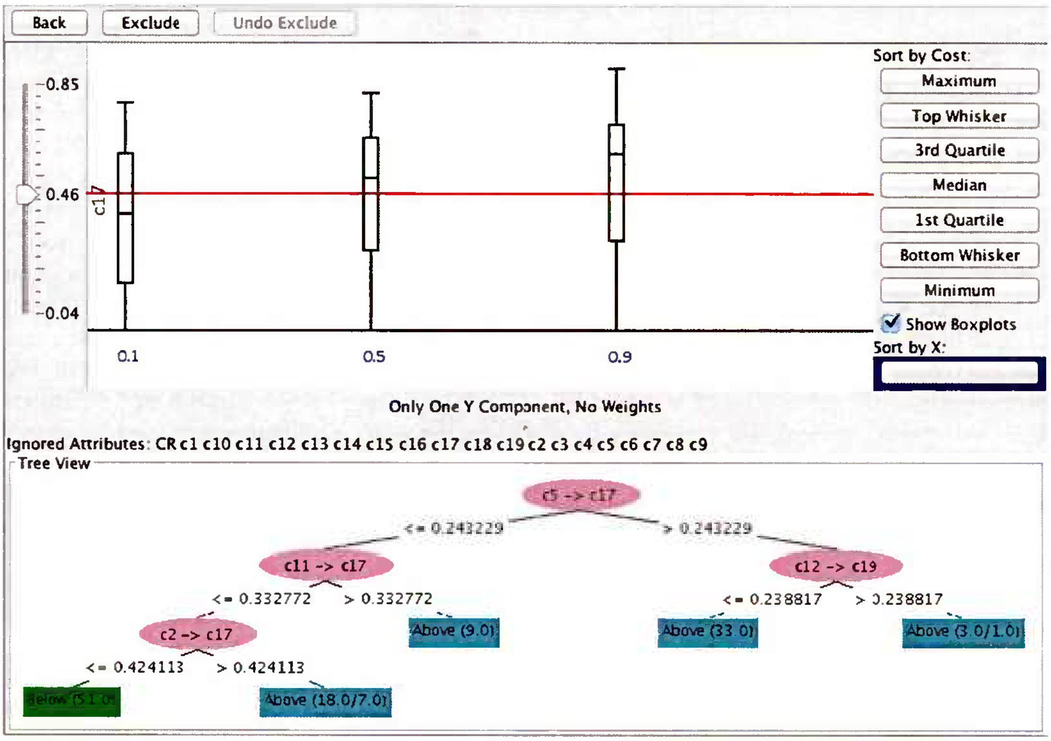


Figure 4. Decision space visualization of the effect of low, medium, and high consumer desire to be green affects the number of electric powered cars consumers purchase

The outcome space can then be represented by a decision space visualization (DSV; Figure 4). In the top portion, the X-axis corresponds to each permutation of a scenario or course-of-action option, and the Y-axis corresponds to the value of an outcome node. For multiple outcome nodes, the Y-axis may include multiple measures for a weighted composite value. This DSV displays the distribution of outcomes for each potential decision option under various plausible conditions. For example, in the case where the Y-axis is cost, and higher cost is bad, options with a low and tight distribution represent more robust options, those which are likely to turn out well under a wide variety of conditions. Options with broader distributions show higher sensitivity to conditions in the model and warrant caution. The tree diagram on the bottom portion of the display, generated using the WEKA package (Hall et al., 2009), represents a hierarchy of underlying interacting characteristics explaining the outcomes above or below the threshold selected in the top portion (the red line at 0.46). The tree is ordered in descending level of influence on the outcomes under inspection. This interactive visualization helps the decision maker understand which nodes in the model are more or less influential on given outcomes in the decision space. By manipulating the horizontal line on the upper diagram to change the threshold of what is considered to be desirable versus undesirable outcomes, this visualization approach allows a decision maker to actually see relationships between options that are otherwise obscured, rather than mentally simulating each one.

We conducted a study to validate the DESIM system in practice. The car-buying model described above was divided into 22 node pairs and the pairwise comparison process described above was administered to 38 workers recruited from Amazon Mechanical Turk, who were paid $4 for their time. Data was collected in two batches. The first 11 individuals responded in 90 minutes after releasing the task. Data collection was closed temporarily to verify the integrity of the incoming data, and then reopened two days later for two hours, collecting another 27 responses. AHP analysis (Harker, 1987) calculated the distributions of edge weights from the respondents' pairwise comparisons. The Java FCM library (De Franciscis, 2014) was used to compute outcome values for the proportion of gas and electric cars consumers purchase (numbers between 0 and 1), for three different scenarios (low, medium, and high consumer desire to be green), with all other input nodes held constant. Consumer desire to be green refers to an individual's preference to make decisions intended to benefit the environment. The FCM was computed for each of the 38 sets of edge weights, resulting in distributions of 38 outcomes for each of the three scenarios, which are shown in the top portion of Figure 4.

RESULTS

In Figure 4, the outcome of interest is node C17, the number of electric cars consumers purchase. Preliminary analysis of the outcomes showed them to be bimodal, so the threshold is set to 0.46 to differentiate the top half from the bottom half of the outcomes for the desire to buy electric cars (the boxplots may be toggled off to show the raw data points, not shown here).

It is clear that as the consumer's desire to be green increases, so does the prediction for the median number of electric cars consumers will purchase, as expected. However, the bottom portion of Figure 4 helps explore the bimodality in the data. The factors under consideration are the various edge weights provided by the crowdsourced population described above. The bottom display has calculated the rules explaining what makes outcomes score above or below the threshold of 0.46, across all three values of the desire to be green. According to the tree display, the most important discriminating factor is the edge between nodes C5 (electric car ownership costs) and C17 (number of electric cars consumers purchase). When an expert rates the weight of this edge greater than 0.24, all outcomes are above the threshold. However, when this edge is rated less than 0.24, the next most influential edge is the relationship between C11 (gas car ownership costs) and C17. When this edge is rated greater than 0.33, the outcomes are above the threshold. Finally, if that edge is rated less than 0.33, it comes down to the expert's rating of the edge between C2 (consumer desire to be green) and C17. Therefore, what explains the outcomes below the threshold are that they are the opinions of experts who believe that all three edges mentioned above have weights below the three indicated tipping points. It also indicates in descending order which of the relationships are most influential and therefore the ones most worthy of attention for decision making. Not shown in this diagram is the ability to select one or more specific outcomes in the top portion of the display, which highlights the corresponding leaves in the tree in the bottom half; the reverse is also possible (Drury, et al., 2012).

CONCLUSION

In summary, the DESIM system elicits mental causal descriptive models from people and transforms them into computer processible causal simulation models, which allows for offloading the simulation-modeling burden from people to the computer. Consequently, by returning choice to a perceptual comprehension process in a decision space visualization, this approach enables decision makers to apply their more powerful visual pattern matching capabilities, rather than their more limited capacities for mental simulation.

Our future work will apply a web-based crowdsourcing system to participatory descriptive model development so that interviews and manual model creation will no longer be necessary. A similar existing system is Scheherezade (Li, Lee-Urban, & Riedl, 2012), a crowdsourcing tool that elicits domain knowledge to create causal narrative models on a given topic, called "plot-graphs." It relies on a structured natural-language processing approach to develop a narrative diagram, represented as a directed acyclic graph. Turkomatic (Kulkarni, Can, & Hartmann, 2012) uses a related aproach to have crowd workers break down a given task into a detailed workflow, making it suited for representing procedural knowledge in the form of a causal model. The contribution of our work, as the first project to use crowdsourced mental (causal descriptive) models that translate into computer-based simulation models, is a streamlined route to model-based decision support tools.

ACKNOWLEDGMENTS

We thank Arthur Liu and Scott Musman for development of the decision space visualizer tool, and Dimitri De Franciscis for support with the Java FCM library. This work was partially supported by the MITRE Corporation Innovation Projects 19MSR080-BA under the Systems Engineering Investment Area and 19MSR083-CA under the Data to Decisions Investment Area.

REFERENCES

Bankes, S. (1993). Exploratory modeling for policy analysis. *Operations Research*, *41*(3), 435–449. doi:10.1287/opre.41.3.435

Caldwell, B. S. (2014). Cognitive challenges to resilience dynamics in managing large-scale event response. *Journal of Cognitive Engineering and Decision Making*, *8*(4), 318–329. doi:10.1177/1555343414546220

Cañas, A. J., Hill, G., Carff, R., Suri, N., Lott, J., & Gómez, G. (2004). CmapTools: A knowledge modeling and sharing environment. In A. J. Cañas, J. D. Novak, & F. M. González (Eds.), *Concept maps: Theory, Methodology, Technology. Proceedings of the First International Conference on Concept Mapping* (pp. 125– 133). Pamplona, Spain: Universidad Pública de Navarra.

Chandrasekaran, B. (2005). *From optimal to robust COAs: Challenges in providing integrated decision support for simulation-based COA planning*. Columbus, OH: Ohio State University, Laboratory for AI Research.

Chandrasekaran, B., & Goldman, M. (2007). Exploring robustness of plans for simulation-based course of action planning: A framework and an example. In *Paper presented at the IEEE Symposium on Computational Intelligence in Multicriteria Decision Making* (pp. 185–192). Honolulu, HI.

Craik, K. J. W. (1943). *The nature of explanation.* Oxford: University Press, Macmillan.

De Franciscis, D. (2014). JFCM: A Java library for fuzzy cognitive maps. In E. I. Papageorgiou (Ed.), *Fuzzy cognitive maps for applied sciences and engineering* (Springer-V., pp. 199–220). Berlin Heidelberg.

Drury, A. J. L., Klein, G. L., Musman, S., Liu, Y., & Pfaff, M. (2012). Requirements for data mining the decision space. In *Proceedings of the 2012 International Command and Control Research and Technology Symposium*. Fairfax, VA.

Gray, S., Gray, S., Cox, L. J., & Henly-Shepard, S. (2013). Mental Modeler: A fuzzy-logic cognitive mapping modeling tool for adaptive environmental management. *2013 46th Hawaii International Conference on System Sciences*, 965–973. doi:10.1109/HICSS.2013.399

Hall, M., Frank, E., Holmes, G., Pfaringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An update. *SIGKDD Explorations*, *11*(1).

Harker, P. T. (1987). Incomplete pairwise comparisons in the analytic hierarchy process. *Mathematical Modelling*, *9*(11), 837–848. doi:10.1016/0270-0255(87)90503-3

Howe, J. (2006, June). The rise of crowdsourcing. Retrieved December 12, 2014, from http://archive.wired.com/wired/archive/14.06/crowds_pr.html

Klein, G. A. (1998). *Sources of power: How people make decisions*. Cambridge, MA: MIT Press.

Kosko, B. (1986). Fuzzy cognitive maps. *International Journal of Man-Machine Studies*, *24*(1), 65–75.

Kulkarni, A., Can, M., & Hartmann, B. (2012). Collaboratively crowdsourcing workflows with Turkomatic. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work -CSCW '12* (pp. 1003–1012). New York, New York, USA: ACM Press. doi:10.1145/2145204.2145354

Li, B., Lee-Urban, S., & Riedl, M. O. (2012). Toward autonomous crowd-powered creation of interactive narratives. In *Eighth Artificial Intelligence and Interactive Digital Entertainment Conference* (pp. 20–25). Stanford University.

Lin, J., Sadeh, N., Amini, S., Lindqvist, J., Hong, J. I., & Zhang, J. (2012). Expectation and purpose. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing -UbiComp '12* (pp. 501–510). New York: ACM Press. doi:10.1145/2370216.2370290

Osei-Bryson, K.-M. (2004). Generating consistent subjective estimates of the magnitudes of causal relationships in fuzzy cognitive maps. *Computers & Operations Research*, *31*(8), 1165–1175. doi:10.1016/S03050548(03)00070-4

Özesmi, U., & Özesmi, S. L. (2004). Ecological models based on people's knowledge: A multi-step fuzzy cognitive mapping approach. *Ecological Modelling*, *176*(1-2), 43–64. doi:10.1016/j.ecolmodel.2003.10.027

Pfaff, M. S., Klein, G. L., Drury, J. L., Moon, S. P., Liu, Y., & Entezari, S. O. (2012). Supporting complex decision making through option awareness. *Journal of Cognitive Engineering and Decision Making*, *7*(2), 155–178. doi:10.1177/1555343412455799

Saaty, T. L. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research, 48*(1), 9–26. doi:10.1016/0377-2217(90)90057-I

Sieck, W. R., Rasmussen, L. J., & Smart, P. R. (2010). Cultural network analysis: A cognitive approach to cultural modeling. In D. Verma (Ed.), *Network Science for Militar Coalition Operations: Information Extraction and Interaction* (pp. 237–255). Hershey, PA: IGI Global.

Surowiecki, J. (2005). *The Wisdom of Crowds*. New York: Anchor Books.

Walker, W. E., Lempert, R. J., & Kwakkel, J. H. (2012). *Deep uncertainty*. Retrieved from http://www.hau.gr/resources/toolip/doc/2012/05/10/deep-uncertainty_warren-e-walker.pdf.

Assessing the Value of Cognitive Task Analysis Reports for Clinical Redesign in Low-Resource Primary Care Practices

Georges POTWOROWSKI[a], Lee GREEN[b]
[a]*University at Albany, Albany, NY, USA*
[b]*University of Alberta, Edmonton, AB, Canada*

ABSTRACT

Thousands of primary care practices face transformational change involving complex, inter-dependent knowledge work. Small, low-resourced practices struggle with the change, and could benefit from CTA. Unlike larger organizations, they lack the change capacity to act on a typical CTA report. We hypothesized that they could use a report that explained how implementing a clinical quality management system (CQMS) would impact their clinical routines, and how adequate their organizational routines were to implement the CQMS. Reports identified deficits in macrocognitive processes, potential consequences of ignoring deficits, and concrete solutions. Two of three clinics made effective changes using this CTA report format. **(100 words)**

KEYWORDS

Low-resource organization; Macrocognition; Cognitive Task Analysis; Health; Change capacity; Implementation.

CORRESPONDING AUTHOR:

Georges Potworowski
1 University Place, Room HPM185
Rensselaer, NY, 12144
gpotwo@albany.edu
(518) 402-0332

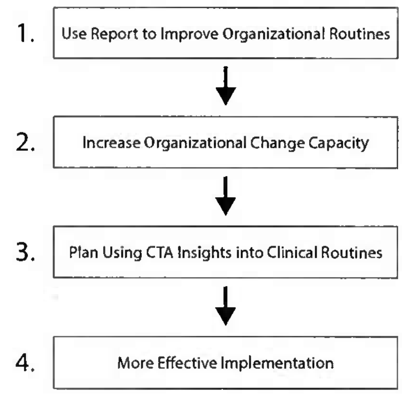
INTRODUCTION

Primary care is undergoing a major transformation in how care is delivered, to the "patient-centered medical home" (PCMH) model ("Patient-Centered Primary Care Collaborative," 2014). This transformation is characterized by a fundamental shift from care that is designed around provider preferences to patient-centric care design. There are multiple domains of care involved ("Patient-Centered Medical Home (PCMH)," 2014), but the two central components of this transformation are a shift from the physician-with-helpers model to intra- and inter-organization team-based care delivery, and an ever-increasing use of technology and information systems to improve the speed, quality, and continuity of care (termed "systems-based care" in the PCMH literature).

The transformation process itself places substantial macrocognitive demands on physicians and staff members, and the new clinical workflows often require extensive reconfigurations of existing macrocognitive processes and functions. Cognitive Task Analysis (CTA) consultation can highlight how key macrocognitive functions and processes would be affected by a given change, and thereby inform practices in their transformation efforts. CTA analysis done for larger, more sophisticated organizations, such as branches of the military and hospitals, often lead to reports that identify insights, opportunities, and issues (AKA "seeds") that the organization can (or should) explore in evaluating their options. Thoughtful consideration of these so-called seeds before acting requires a degree of change capacity and organizational slack that these practices simply do not possess.

To address this capacity issue, we hypothesized that a two-part CTA analysis leading to a more concrete, prescriptive, and supportive report would prove usable by these practices. The first analysis focused on how implementing a clinical quality management system (CQMS), a form of health information technology, would impact their existing clinical routines. The second analysis focused on the adequacy of their organizational routines to implement the CQMS. The resulting reports were concrete, prescriptive, and supportive by first identifying specific macrocognitive deficits in clinical and organizational routines relevant to CQMS to inform clinics. For each deficit, the report explained the potential consequences of not addressing it to motivate them, and offered concrete solutions for addressing it to guide them. The objective (see fig 1) was to have practices use the report to improve their organizational routines, which would increase their change capacity. This would improve their implementation planning by allowing them to use the CTA insights into their clinical routines. Doing so would then increase their chances of successful implementation.

Fig 1. Logic Model Connecting Two CTA Analyses



METHODS

We prepared interview guides and probe questions for Task Diagram and Team Knowledge Audit (Potworowski & Green, 2013) through multiple iterations over a period of several months by the full team. The probes were informed by the literature on organizational routines (Becker, 2004; Greenhalgh, 2008), LG's experience designing the CQMS, LG's previous experience with CTA in primary care, LG's expertise as a primary care physician, GP's expetise in cognitive and organizational psychology, discussions with colleagues expert in IT and sociotechnical systems, and trial runs with non-study practices.

Each site's participation began with an organizational meeting. The project team visited the site, presented a detailed discussion of the project, and worked out a launch plan and timetable with the site. The CQMS sales team then provided their standard commercial demonstration and introductory training presentation to the site.

The research team conducted the interviews at the clinics in the typical CTA process—by a pair of interviewers (lead and second). Multiple informants covering all roles in the clinic were interviewed. Detailed Task Diagrams and Team Audits were conducted with each interviewee independently, then integrated in subsequent analysis, in order to identify dispersed vs. distributed knowledge and understand the degree of commonality of mental models among team members. Notes were taken during each interview, and the interviews were recorded and transcribed for later analysis. The team also collected extensive field notes on the physical environment, observations of interactions, impressions of the organizational climate and culture, and work activities observed.

In our first round of analysis for each practice, we studied our data to understand the practice's clinical and organizational routines, and produced a CTA report as a guide to implement the HIT. Analysis started with members of the project team performing an initial round of immersion crystallization and individual proposals for emergent themes. They then met as a team weekly to refine the coding scheme, which resulted in six macrocognitive processes: Decision Making, Sensemaking and Learning, Planning and Re-planning, Problem Detection and Monitoring, Managing the Unknown and Unexpected, and Coordinating. The team then coded the transcripts, and reconciled coding. Further weekly meetings were devoted to generating a set of specific action recommendations and information points, and preparing a detailed custom report (approximately 20 pages) for each practice. Each report highlighted tacit and dispersed knowledge crucial to their clinical routines, areas that the CQMS would affect and how, decision points, likely failure points, current information handling, and workarounds. It summarized similar features in their organizational routines. The report's recommendations for work routines covered recommended changes in workflow and information handling, knowledge sharing, mitigation of failure points, and obviating workarounds. For organizational routines, recommendations focused on specific risks to, or changes needed to support, implementing the workflow changes recommended.

After conducting the first analysis and preparing the CTA report, we then returned to the site to present the report and discuss our recommendations. In the first clinic, staff came an went during the presentation. In the second and thrid clinics, the entire staff of the clinic was present and discussed. Further detailed discussions about the recommendations were held subsequently with leaders responsible for the implementation. At that point the CQMS vendor representatives conducted their standard commercial software installation, hands-on training, and launch.

We followed up by telephone conference with the clinic leadership after launch, and at the second and third sites made follow-up site visits on several occasions up to 18 months later. Detailed field notes were kept on the teleconferences and at the site visits, and additional CTA interviews and direct observations were conducted at the site visits. These data were analyzed in the same fashion as the first round, again in weekly team meetings. This second round analysis focused on applying the macrocognition framework to understand the impact of the CTA consultation report, the outcome of the implementation, and clinic's management of change at the clinical and organizational levels. Several weeks later, we visit the practice to learn what use they have made of the report, and gather their input for further improvement of our method.

RESULTS

CTA methods were readily applied for both understanding the teams' workflow routines and evaluating their change routines. The process was very well accepted by all 3 sites, with many comments about just going through the interviews being helpful to the practice, even before results were received.

In the first clinic, staff participation in the report delivery meeting was limited as patient visits were ongoing. Later follow-up and debrief interviews with clinic management and the senior clinician on staff indicated that while they had positively received our report and agreed that our advice/suggestions were valuable, they failed to follow the majority of them in the implementation process. For example, we suggested adjusting the patient visit schedule to accommodate the extra time needed for data entry and workflow adjustment during the roll out period, but the clinic chose to attempt to continue operations at full capacity. They did launch the CQMS on schedule, and ran it for over 3 months. Ultimately, their lack of planning and poor change routines ultimately meant implementation problems were not solved and the clinic discontinued their use of the CQMS.

In the second clinic, they introduced our team and project to their staff and providers at an all-hands meeting of personnel convened from all their sites. Later, when we presented our report, they convened site leaders and staff from across their dispersed sites to receive our report and discuss it. This clinic began to implement some of the recommended changes, but ultimately faced human resource and financial problems that prevented them from launching the CQMS and ultimately led to the clinic's closure.

The third clinic in our study had experienced a previous failed attempt at implementing the same CQMS product. Initial CTA interviews revealed a basic quality improvement skill set that could be leveraged for change management to improve their existing but limited planning, replanning and monitoring skills. Clinic leadership was very receptive to our research team's involvement with their committees responsible for quality improvement and CQMS implementation. They closed the clinic and had all staff present when we delivered our CTA report. Staff at all levels participated actively, asking questions and discussing action plans. They included us in their planning, communications, and meetings, and actively sought our consultation and support. In followup interviews, they reported executing a number of the strategies presented in our CTA report, in particular those that involved more formal problem detection and monitoring, replanning, and coordination. They invested the staff time, meetings, and resources recommended. Over the course of our involvement with them, they substantially improved their problem detection and monitoring, replanning, and sensemaking processes, becoming steadily less dependent upon our support. In the end, they succeeded in implementing the CQMS.

DISCUSSION

CTA was readily applied and sustained, acceptable to practices (the process was acceptable even where its results were not applied), and plausibly effective in the safety net primary care setting. More broadly, the macrocognitive functions that CTA evaluates are of great importance for practices, in this period of rapid transformational change in primary care. This project demonstrated that CTA can identify specific deficits in those functions, and CTA-guided intervention can improve a team's skills at those functions, in primary care as it has in other areas of knowledge work. That suggests that the large body of literature on CTA does generalize to primary care, and could add significant value to practice facilitation or change management efforts.

CONCLUSION

In this project we have demonstrated that a two-part CTA analysis leading to a more concrete, prescriptive, and supportive report can a) help small, low resourced primary care practices improve their change routines, which then b) allowed them to avail themselves of CTA insights into how to improve their clinical routines.

ACKNOWLEDGMENTS

We thank the practices that participated in these projects, and the Altarum Institute staff for collecting and analysing data. We thank the Agency for Healthcare Research and Quality (grant R18 HS018170-01) for its funding support.

REFERENCES

Becker, M. C. (2004). Organizational routines: a review of the literature. *Industrial and Corporate Change*, *13*(4), 643-678.

Crandall, B., Klein, G. A., & Hoffman, R. R. (2006). *Working Minds: A Practitioner's Guide To Cognitive Task Analysis*. Cambridge, Mass.: MIT Press.

Greenhalgh, T. (2008). Role of routines in collaborative work in healthcare organisations. *BMJ*, *337*, a2448-a2448. doi:10.1136/bmj.a2448.

Patient-Centered Medical Home (PCMH). Retrieved May 4, 2014, from http://www.ncqa.org/Programs/Recognition/PatientCenteredMedicalHomePCMH.aspx

Patient-Centered Primary Care Collaborative. Retrieved April 6, 2014, from http://www.pcpcc.org/

Potworowski, G., & Green, L. A. (2013, February). Cognitive Task Analysis: Methods to Improve Patient-Centered Medical Home Models by Understanding and Leveraging its Knowledge Work. Agency for Healthcare Research and Quality, Publication No. 13-0023-EF.

Expertise and decision making in real estate appraisal: results from a naturalistic study

Olga PREVEDEN
University of Vienna, a0849519@univie.ac.at

ABSTRACT
This project presents the results of the application of Applied Cognitive Task Analysis methodology to the domain of real estate appraisal. The research into appraiser decision making for the past 20 years has been dominated by the behavioral tradition based on rationalistic models. Naturalistic approach allows an alternative view stressing the role of expert cognition in decision making. The data has been collected in interviews with 10 Austrian appraisers and analyzed by means of thematic analysis. The analysis of qualitative data supports the assumption that real estate appraisal is more an art than a science.

KEYWORDS
Real estate appraisal, expertise, appraiser decision making; ACTA.

INTRODUCTION

The problem of real estate appraisal

This paper presents a part of a research project that focuses on the decision making of Austrian real estate appraisers and explores the requirements for the decision support in this field. Real property is the main type of the assets of companies and whole economies and therefore it is vital to know the property's value. However, definition of the property's value is not a trivial task: there are multiple types of values calculated for different purposes. Most often real estate appraisers are asked to calculate *market value* of a property that is an estimation of the most likely transaction price in the open market. Such estimation is difficult as a number of physical characteristics of real property provide for its low efficiency: real assets are immovable, durable, scarce, and very heterogeneous. As a consequence, real estate market is characterized by high segmentation and the conditions of uncertainty stemming from limited information, low market liquidity and low market stability in certain phases of the business cycle. These conditions, on the one hand, provide for the demand for the appraisal services, but, on the other hand, constitute the main challenge of the appraiser's profession. The task of the appraiser is to deal with this uncertainty (Crosby, 2000; French & Mallinson, 2000).

Appraisal uncertainty and accuracy

There have been many attempts to represent the uncertainty and to measure the related construct of appraisal accuracy, especially since the question of appraisal accuracy had been raised for debates in the Mallinson report in 1994. No adequate representation of uncertainty has been found and there have been ongoing debates about an adequate measure of appraisal accuracy (valuation error vs. valuation variation) (Crosby, 2000; Kucharska-Sastiak, 2013). The results of empirical studies on appraisal accuracy are not uniform: the studies with positive results prevail numerically, but there are also studies seriously questioning the quality of appraisal services.

Appraisal procedure

Facing the lack of an adequate accuracy measure, professional organizations focused on regulation of the appraisal process. There are three generally accepted appraisal approaches: income capitalization approach, sales comparison approach and cost approach. The sequences of steps for each approach are prescribed on the national level and are binding. The approach to study decision making in real estate appraisal has long been dominated by the rationalistic tradition coming from finance, e.g. DCF approach is widely used for income properties transacted internationally. But the norms leave a lot of decision space for the appraiser in setting the input parameters for the calculation and they have a crucial impact on the valuation results.

Research on appraiser decision making

Behavioral research tradition found its way into the study of appraisal decision making relatively late. The view on appraisal as information processing task under uncertainty originates from Richard Ratcliff (1972). Only eighteen years later Diaz (1990) applied human problem solving approach from cognitive psychology in practice. Since then behavioral approach has dominated the field of real estate appraisal. Behavioral studies cover four main topics: 1) deviations from normative models 2) comparable sales selection 3) the use of heuristics and biases, and 4) the role of client pressure (Diaz 1999). A conclusion from these studies is uniform: appraisers deviate from the prescribed normative procedure, they use information cues selectively and are viable to different biases. Typically for behavioral research, deviations from normative models are seen as cognitive risks (Wofford et al., 2011). An alternative claim was formulated by Hardin (1999) who proposed to test whether expert heuristics are always functional and if there is a place for expertise in real estate appraisal. But this claim has not been addressed ever since. Recently in the literature on appraisal accuracy expertise has been addressed again with the claim that expert appraisers are able to reduce the uncertainty of the input data through their way of information processing (Kucharska-Sastiak 2013).

STUDY

Purpose of the research and research questions

Real estate appraisal is a complex problem solving conducted in complex conditions. The success of performing an appraisal task depends a lot on the knowledge and expertise of the appraiser. No studies have examined expertise in the domain and an understanding of how appraisers make decision is still lacking. Several research questions have been formulated: What makes real estate appraisal task complex? What constitutes expertise in real estate appraisal? What are the main cognitive activities of the appraisers? What information cues are important?

Research method

Given the characteristics of the domain of real estate appraisal it has been assumed that the application of Naturalistic Decision Making theory and its methods would allow new insights in the domain. This paper presents the results of application of Applied Cognitive Task Analysis methodology (ACTA) to study appraisers' expertise and decision making. ACTA methodology (Militello et al., 1997) has been developed based on the findings from the psychological research on expertise and is targeted to study expertise and cognitive activities in a particular domain. The guidelines for semi-structured interviews have been prepared on the basis of ACTA methodology with regards to the research questions (Militello et al., 1997). Stage 1 (Task diagram) and Stage 2 (Knowledge Audit) were completed in the full extent. The probes for Knowledge Audit included Past & Future, Big Picture, Noticing, Job Smarts, Opportunities/ Improvising, Self-monitoring, Anomalies, System difficulties and Scenario from Hell. Stage 3 (Simulation Case) has been carried out with a real-life case without financial figures due to time constraints and voluntary basis of the participation; the participants have been queried how they would proceed and what would be the main possible difficulties. Each interview took on average 1.5 hours.

Research participants

ACTA interviews were conducted with Austrian real estate appraisers in German language. The participants have been selected on the basis of their membership in national and international membership organizations. Current paper is based on the results of 10 interviews. The average age of interview participants comprised 47.2 years; their track record on appraisal equaled on average 13 years (ranged from 3 to 20 years) and each participant appraised on average 142 properties per year (ranged from 50 to 500 properties).

Data analysis

The interviews were tape recorded and transcribed for the analysis. QTA Miner software was used to organize and analyze the qualitative data. Thematic coding methodology was applied to code the ACTA interview data. Themes have been formulated based on the prior NDM research in the related fields (finance, management) and based on real estate appraisal literature. The coding categories included: expertise (with subcategories Big Picture, expert intuition, interest, schemas, metacognition, noticing, tricks of trade), expert knowledge (previous projects, education, experience, general knowledge, personal contacts, market knowledge, principles, soft skills), cognitive demands (demand, Why difficult?), decision (alternatives, ambiguity, liability, mental simulation, sensemaking), information (cues, key indicators, sources of data), comparables selection (adjustments, criteria, yield), industry specifics (industry standards, market conditions, property uniqueness).

RESULTS AND DISCUSSION

ACTA Stage 1: Task diagrams

The appraisal process is highly standardized (Figure 1). When performing an appraisal task, an appraiser creates a "story" of the property. He starts by looking through the available documents to imagine what kind of property it can be and gains his first impression of it. He proceeds with the inspection of the property and gains his second impression of the property. An additional screening of documents maybe be needed. At the next step, comparables are selected to define the input parameters for the calcualtion. The selection of inputs parameters has been defined as the most cognitively demanding task performed by the apprasiers that can be attributed either to market research or to calculation stage.

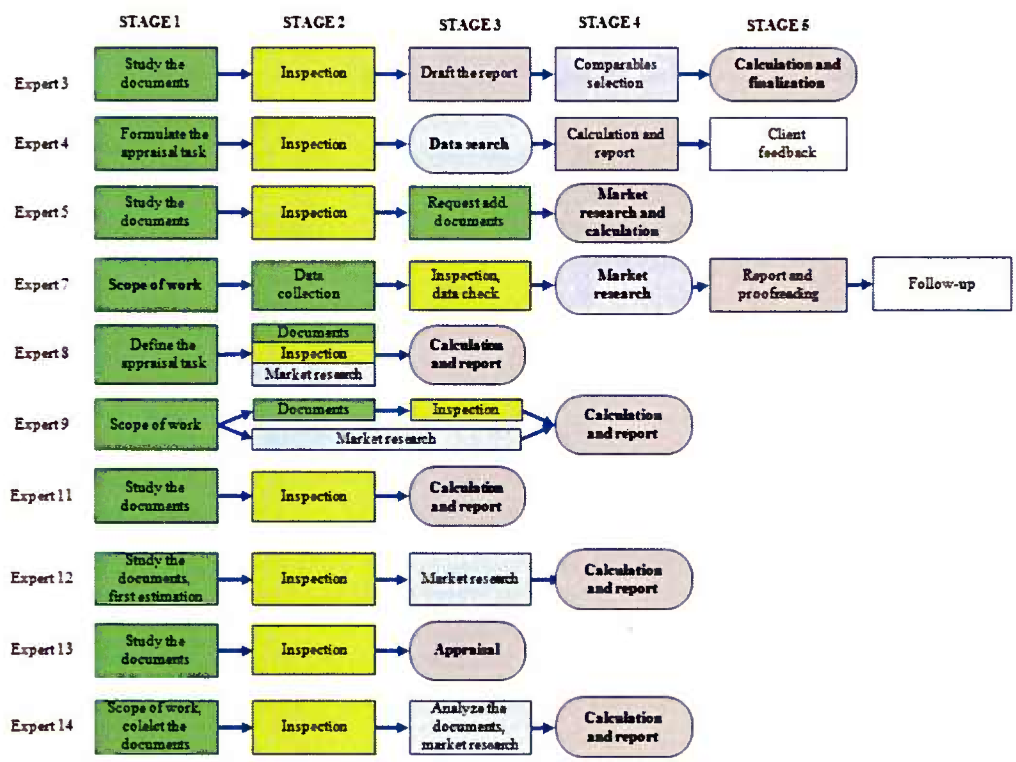


Figure 1: ACTA Stage 1 results– Task diagrams of participants

ACTA Stage 2: Knowledge Audit results

Expertise of real estate appraisers

Real estate appraisal is an expertise-dominated domain. For all interview participants it is important to cooperate with renowned colleagues, whom their know personally. The necessary conditions for expertise are professional education and several years of professional experience. A logically written appraisal report is more important to judge the quality of the appraiser services than the market value itself. In this respect real estate appraisal can be compared to the domains of legal services and business intelligence. Appraisers often refer to their own experience and market knowledge as an information cue. It is important to know the market (rents, yields), to know the buildings, and relevant transactions. If market knowledge is lacking the typical way is to contact other market participants.

Real estate appraisers demonstrate a number of typical traits of experts: they are self-confident, they can detect anomalies, they are better in combination of information cues. Several participants also stressed their professional intuition or a feeling of just knowing the value. Having this challenging job was mentioned as an important factor that drives professional development.

It comes out that knowledge and experience are more crucial in real estate appraisal then specific procedures. Further research can be conducted with the purpose to find out how the appraiser's knowledge is structured and organized.

Cognitive processes and demands

The process of setting the input parameters is the heart of real estate appraisal. Real estate appraisers think in terms of market segments and comparables. Market knowledge of an appraiser can be represented as a map characterized by rents and yields. The task of the appraiser is to place the property in question among other properties on this map.

Cognitive Demand Tables have been constructed and 57 cognitive demands have been identified, all of which can be attributed to definition of the relevant market segment, comparables selection or setting of the input parameters for the calculation. The process of setting the parameters is cyclical: the parameters are adapted until there is a feeling of being right. Dependent on information quality and experience of the appraiser, more or less steps than it is normatively prescribed, can be required to decide on the parameters.

Not being explicitly addressed the topic of probabilities was not mentioned by the appraisers. In combination with the finding that the appraisal report is more important than the valuation figure itself, this confirms the well-known proposition that appraisal is more an art than a science and moves its focus away from quantitative approaches.

Comparables selection

The general concern of Austrian appraisers is the lack of relevant data. Assignments where no comparables exist or where the data is not publicly available were named to be the most challenging ones. Ideally the appraisers wish to have 5 or 6 comparables, but mostly they have to come out with just one or two. A decisive step in comparables selection is price adjustment, but the appraisers agree that to develop a weighting rule for different attributes or some scale of price adjustment would be a greater endevour than to appraise a single property without such tools or to train a new colleague how to perform an appraisal. How the weights are attributed constitutes important knoweldge of appraisers and it is the task of further research to study how this knowledge is formed and used.

Complexity of real estate appraisal

Besides the lack of information other factors contributing to the complexity of appraisal assignments were: mixed properties (mentioned by 4 persons), legal problems (9 persons), tenant problems (single tenant) (4 persons), rent problems (underrent/ overrent) (4 persons), properties requiring maintanence (4 persons), and properties in a second-order location (B class) (2 persons). The combination of different factors increases the complexity of an assignment as more cues have to be combined and the cognitive load is still higher.

CONCLUSION

This research applied Naturalistic Decision Making approach to study the expertise and decision making of real estate appraisers. The results of ACTA analysis demonstrate that appraisers' decision making is fixed on the selection of input parameters for the calculation and it is more a qualitative endeavor than it is typically considered in appraisal literature. Expert cognition is important for the success of appraisal services and is based rather on specific knowledge and experience than on specific procedures.

REFERENCES

Crandall, B., Klein, G. & Hoffman, R. H. (2006) Working minds: A practitioner's guide to CTA, MIT Press, London.

Crosby, N. (2000) Valuation accuracy, variation and bias in the context of standards and expectations, *Journal of Property Investment & Finance*, 18, 2, 130-161.

Diaz, J., III. (1990) How Appraisers Do Their Work: A Test of the Appraisal Process and the Development of a Descriptive Model, *Journal of Real Estate Research*, 1990, 5:1, 1–15.

Diaz, J. (1999) The first decade of behavioral research in the discipline of property, *Journal of Property Investment & Finance*, 17, 4, 326 - 336.

Diaz, J., III (2010) Disrobing beautiful people: an introduction to the special issue of behavioral real estate research, *Journal of Property Research*, 27, 3, 203-206

French, N. & Mallinson, M. (2000) Uncertainty In Property Valuation. The Nature and Relevance of uncertainty and how it might be measured and reported, *Journal of Property and Finance*, 18, 1, 13-32.

Hardin, W., III (1999) (1999) Behavioral research into heuristics and bias as an academic pursuit: Lessons from other disciplines and implications for real estate, *Journal of Property Investment & Finance*, 17, 4, 333 – 352.

Kucharska-Stasiak E., (2013), Uncertainty of property valuation as a subject of academic research, *Real Estate Management and Valuation*, 21, 4, 17-25.

Militello, L.G. & Hutton, R.J.B. (1998) Applied cognitive task analysis: A practitioner's toolkit for understanding cognitive task demands, *Ergonomics*, 41, 11, 1618-1641.

Ratcliff, R.U., (1972) Is there a "New School" of Appraisal Thought? *The Appraisal Journal*, 40, 522-528.

Wofford, L., Troilo,, M. & Dorchester, A. (2011) Real estate valuation, cognitive risk, and translational research, *Journal of Property Investment & Finance*, 29, 4/5, 372 – 383.

Positioning RPD in the Decision Making Matrix

Agoston RESTAS
National University of Public Service, Budapest, Hungary

ABSTRACT
Introduction: This paper provides a decision making matrix, a unique division of the different type of decision theories. Method: Author used his own practical experience in the field of emergency management, carried out logical explanation and used graphic solution. Results and discussion: based on temporal impacts of decision in the future and the time spent on them author created a simple decision making matrix, in which 4 fields can be found. Each field contains a characteristic decision type, i.e. *classic, bureaucratic, routine* and *recognition-primed decisions* (RPD).

KEYWORDS
Common ground, modelling, decision matrix, emergency management, RPD

INTRODUCTION

There are many ways to classify decision theories. In many cases researchers used two distinct methods. One of them defines the principles and rules based on which the decision maker has to reach the final result. These methods belong to the so-called normative models. The other method basically focuses on the decision-maker as a thinker and emotionally charged person and the formation process of decisions made by him. Methods in this group are called descriptive models. Naturally there are also other aspects of classification. Different levels of decision making can be found almost in all organizations; it means strategic, operational and tactical levels. Moreover people often say for administrative work made by authorities as bureaucratic decision. In this jungle author as an emergency manager intends to position his practice, where recognition primed decision as a demonstrative part of naturalistic decision making is often used.

DECISION MAKING MATRIX

Decision Types Based on their Temporal Impacts and the Time Spent on Them

To explain and understand the essence of special decision-making mechanisms, what emergency managers, e.g. fire fighters often use, author creates a unique *matrix*. In this matrix, author regards the *magnitude* of *time spent on decisions* and the *temporal impact of decisions, its "weight"* as determining characteristics.

When establishing the matrix, author set the requirement that it may not infringe the regularities of analogical decisions, nonetheless, it is able to demonstrate the structure of our decisions in a way different from the tradition ones, so the unique decision-making mechanism of those in emergencies is included with emphasis.

Decision from Strategic to Tactical Level

The weight of the decisions of managers, paired with the division according to its time horizon can be also found in the work of many experts (e.g.: Radford, 1988; Molnar, 2003; Kelly, 2011). In the center of division, organizations with different structures stand, where "heavy-weight" decisions, i.e. *strategic decisions* are made by *senior managers*, "middle-weight" decisions, i.e. *operational decisions* by mid-level managers, "light-weight" decisions, i.e. *tactical decisions* by low-level managers. The time horizon of decisions means a *long-*, *mid-* and *short-term* division. Molnar, in his summary, does not directly link strategic decisions with the long-term time horizon, furthermore, the tactical one with the mid-term one, and the operational one with the short-term one, however, logically, this content is unambiguously in the background. In the scope of management and decision theory, this concept can be justified through the works of a multitude of different authors (e.g., Kindler, 1991; Greco, 2005; Bakacsi, 1996), thus author regards it as generally accepted.

With the illustration of the division, the interrelation between and hierarchy of decisions can be well seen at *Figure 1*. Decisions only occupy a part of the fields defined by the coordinate axes, and based on the logic of division, the "empty" parts do not even exist. "Heavy-weight" decisions cannot be made in a short time, and the weight of decisions, made on operational level, may only be low. This type of division is certainly not so strong with authors preferring this method, but its inner core clearly points in this direction.

The above approach, in author's opinion, has a view on a decision as an end-product from inside an organization and not as an active link with the partner or the environment. Looking at it from outside, the impact and success of decisions, in author's opinion, can be completely different. To obtain a license from the authorities it is worth mere yes for the client a, while inside an organization, it can obviously be evaluated in a different way, however, the strategic decisions of firms in relation to partners can also be regarded as strategic determination.

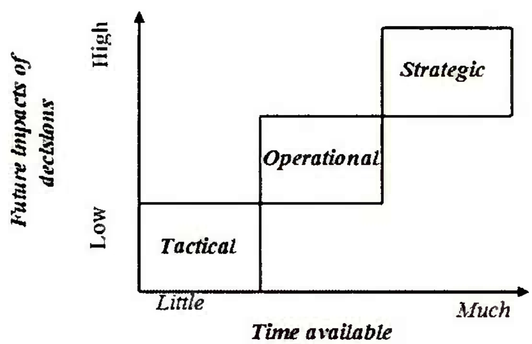


Figure 1 – Relationship between strategic, operational and tactical decisions depending on the time available and the future implications. Source: Author

We know of the classic models described in the previous chapter that the stakeholders of business life use them to achieve their long-term success, mainly strategic objectives. Strategic objectives obviously greatly influence the long-term activities of actors of business life, so, they can be regarded, based on their future impact, as significant, "heavy-weight" objectives. To do so, decision-makers have enough time, compared to the interpretation domain of the concept of emergency, defined in this article, by magnitudes more time.

Simplifying the Decisions

If the significance of the serious impacts of decisions is taken, author assumes that there must be, on the contrary, a decision with a "weight", whose impacts are considerably lower. We all practice them daily, regarding them as *routine-like*; based on this, it is named *routine decisions*. Another well-known feature of routine decisions is that not only are their future impacts scarce, but also we only spend a little time to make them; due their automatism, we practically do not even notice them. Despite of this fact, this decision type should not be neglected, since our everyday actions are mainly based on them (Betsch & Haberstroh, 2005; Ribarszki, 1999).

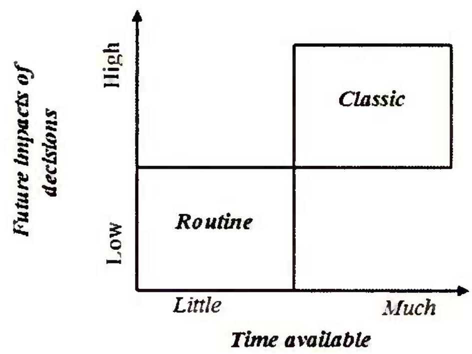


Figure 2 – Relationship between classic and routine decisions depending on the time available and the future implications. Source: Author.

Regarding the interrelation between classic and routine decisions author ascertains that they are converse as far as their future impacts and the time spent on them; the former has significant impact and long time, the latter has scarce impact and short time. The above are illustrated in a coordinate system in *Figure 2*.

Complementing Empty Fields
Following the above train of thought, logically, the question arises whether the unfilled parts in the coordinate system can be filled, i.e. a relatively low-importance decision paired with long decision-making time and its opposite, significant future impact paired with short decision-making time, from the aspect of decision-making procedures.

Author divided the sides of the matrix, i.e. the axes in the simplest way: in the case of time, *little-much*, in the case of the impacts of decisions, *low-high* values. Thus, the matrix gives four fields (*Figure 3*), to which author uses the following names: *classic, bureaucratic*, *routine* and *recognition-primed decisions*. The values of classic and routine decisions, based on the above, have already been defined: in the previous case both values are high, in the latter they are low. The values of the two new fields are contradictory: in the case of bureaucratic decisions, their future impacts are *low*, the time that may be spent on them is *much*. With recognition-primed decision, the situation is opposite: the extent of impact is *high*, the time that may be spent on it is of *little* value.

Thus, the fields of the matrix have been filled, however, it is necessary to review what their content actually means.

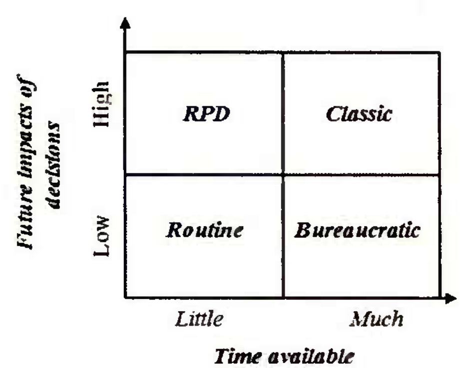


Figure 3 – Decision matrix in relation to the time available and future impacts. Source: author

FEATURES OF THE MATRIX'S FIELDS

Classic Decisions

Researchers of decision theory study this decision-making mechanism in the widest ranges, thus, the different literatures this field of decision theory is the most wholly described (Belton & Stewart, 2002; Paprika-Zoltay, 2002). The characteristics of the field are that decision-making has on both exes "high" values. The action as a result of the decision has a significant future impact. In order to make this decision, careful considerations are necessary, which can only be done with sufficient time spent on it. It means that from the time of recognizing the problematic situation until specific decisions days, weeks or perhaps months may be available. It facilitates the decision-maker to collect information, analyze it, create options based on the results, to modify and compare them by introducing new conditions, or perhaps completely exclude certain options. Options that may bring the best results for the decision-maker, based on the elaboration of information and conditions available in a given time, will be implemented.

The above steps can, of course, be put in another form. Not specifying this decision field any longer, author concludes that we are dealing with a long-term quest, allowing the development of several variations, to which author uses the name: classic decision-making.

During the operation of law enforcement agencies, including the area of disaster management and the fire service, it is the dominant form of decision-making. Each manager is to apply it at different levels, obviously for the sake of facilitating and ensuring long-term efficient operation. A chief fire officer, based on previous intervention statistics and depending on the probable future vulnerabilities of the area under his responsibility, makes an effort to replace equipment due to depreciation, to purchase new equipment, to increase the staff or regroup. These efforts, naturally, are often in contradiction to the will of the senior management, primarily not due to professional disagreements, but budget restrictions (Restas, 2011). The latter depends on the country's fiscal situation, the size of the amounts to be spent on fire protection.

Bureaucratic Decisions

It can be observed as a typical decision-making process at a bureaucratic organization like government organizations or different authorities. The field is characterized by the fact that the problem's weight is *low*, while the time spent on a solution, represents a ***high*** value. The operating mechanisms of these organizations are analyzed by sociology, more specifically by organizational sociology. The essence of the background is decision is not to reach an individual solution, taking into account the characteristics of the given problem, but to prepare a template, aligned to the operational mechanism of the organization and easily manageable. Its simplest example can be the forms and questionnaires of authorities.

Not underestimating even by chance the work performed by such an organization, however, author concludes that from the aspect of decisions, the activities of bureaucratic organizations can best be compared to compliance, after comparison. Specifically, a comparison of the problem's contents takes place with the provisions of an existing sample (mostly legislation), which usually requires a *yes-no* elementary decision, without variations. The organization usually has restricted time, but at least days to make this decision.

In the field of law enforcement agencies, there are also many examples of the above decision-making mechanism. For instance, fire service as a public authority and a professional authority, manages requests submitted to it according to the national acts on the rules of public administration procedure. In its competence, it compares the issues submitted in the requests (such as the establishment and use of buildings) with the relevant legislation in force, and agrees to the decision (authorizes it) or not. The decision-maker does not change the subject of a request in case of non-compliance, it does not recommend or give advice. The simplistic outcome of their decision is the communication to the applicant of a *yes – no* variation. The above are, of course, very simplified descriptions of the process, and the result is similar during the functioning of any other authority (e.g. police, local government).

Routine Decisions

Small actions of daily life are based on this decision-making mechanism. The field's characteristic is that both values of the factors of the matrix are low. This is exactly what individuals need to take to tackle the constantly repeating moments of everyday life not to constitute a decision problem. Many times, it is a subconscious set of activities, whose deeper examination is covered by psychology. Since it is a rerun of identical activities, the brain will automatically give orders to implement it, without committing substantial capacity.

It belongs to the essence of the fact of the decision that basic problems are solved here, to which previously there was the same or similar response. So, by recalling, a process that has already occurred will be repeated. As a result of constant repetitions, one of the characteristic features of decision is the effectiveness of automatism, that is, the time spent on decision manifests itself in its minimum requirement.

RPD – Recognition-Primed Decisions

The field is characterized by the fact that decisions drawing behind serious consequences shall be made in a relatively short time. Classic decision-making mechanism, already discussed, due to the shortage of time, is practically useless, in some cases it may be even dangerous (Klein, 1989).

The comforting weightlessness of routine decisions, by the very nature of the problem, clearly cannot receive a role. The fact that this is a typical decision-making model, is crystallized as a result of a number of observations. It was observed during a military exercise that commanders made the vast majority of their decisions in less than 1 minute. The number of decisions made in more than five minutes was rather scarce. During another survey, involving chief fire officers with over 20 years of practice, having studied 450 decisions of a total of 150 experienced decision-makers, they ascertained that 85% of decisions were made within one minute. They drew the consequence that, different from the analyzing and evaluating thinking, it is a typical decision-making procedure, which they called recognition-primed decision (Klein, 1989). This procedure is the typical decision-making model of professional managers in emergencies, like firefighting (rescue operations) managers, police officer in criminal action, emergency surgery, pilot, etc.

CONCLUSION

Based on author's assumption, the mechanism of decisions can be divided in a way that ensures the equivalence of emergency decision-making. To justify this hypothesis, author created a decision matrix, in which he took as a basis the future impact of decisions and the time spent on it; thus, we received 4 fields. Each field contains a characteristic decision type, i.e. *classic*, *bureaucratic, routine* and *recognition-primed decisions*. The significance of the division lies in the fact that, by doing so, the decision mechanism of emergency decision-makers receives an equivalent decision position from the periphery of mechanisms studied so far.

REFERENCES

Bakacsi, Gy. & Bokor, A. (1996) Szervezeti magatartás és vezetés; (Ed.) KJK_KERSZÖV Jogi és Üzleti Kiadó Kft. Budapest, ISBN 963 224 496 6

Belton, V. & Stewart, T. (2002). Multiple Criteria Decision Analysis, An Integrated Approach. Springer, Vol. XIX, ISBN 978-1-4615-1495-4

Betsch, T. & Haberstroh, S. (2005). The routines of decision making. Psychology Press, Taylor & Francis Group, New York, NY10017, ISBN13: 978-0-8058-4613-3

Dobak, M. et al. (1996). Szervezeti formák és vezetés. Közgazd és Jogi Kiadó, Budapest. ISBN 963 222 972 X

Greco, S. (2005). Multiple Criteria Decision Analysis: State of the Art Surveys. Springer, International Series in Operations Research & Management Science, XXXVI. Vol. 78, ISBN 978-0-387-23081-8

Kelly, M. (2011). Organisational Decisions and Systems. IT Applications Theory Slideshows, Vcet.com, 2011.

Kindler, J. (1991) Fejezetek a döntéselméletből; Aula Kiadó, Budapest, 963 10 1830 X

Klein, G.A. (1989): Strategies of decision making , Military Rewiev, No.5.

Molnár, M. (2003) A vezetői döntés (Chapter III.); In. Vezetéselméleti ismeretek (szerk.: Molnár, M.), Főiskolai jegyzet, Rejtjel Kiadó, Budapest, Humgary, p. 82

Radford, K.J. (1988). Strategic and Tactical Decisions. Springer, ISBN 978-1-4613-8815-9

Restas, A. (2011) An Approach for Measuring the Economic Efficiency of Aerial Fire Fighting; Presentation, Wildfire2011 The 5th International Wildland Fire Conference, Sun City, South Africa, 9-13 May 2011.

Ribarszki, I. (1999) Döntéspszichológia (Decision psychology), Zrínyi Miklós Nemzetvédelmi Egyetem, Jegyzet, Budapest, Hungary

Tari, E. (2004): Max Weber bürokrácia-tanának szervezet elméleti jelentősége korlátja, pp. 24-27; In: Szöveggyűjtemény a Szervezet és vezetéselmélet tárgyhoz Vezetéstudományi intézet, Budapest, Hungary

Zoltayne-Paprika, Z. (2002) Döntéselmélet; Alinea Kiadó, Budapest, Hungary, ISBN 9638630612

Externalizing Planning Constraints for More Effective Joint Human -Automation Planning

Ronald SCOTT[a], Emilie ROTH[b], Beth DEPASS[a] and Jeffrey WAMPLER[c]
[a]*Raytheon BBN Technologies*
[b]*Roth Cognitive Engineering*
[c]*Air Force Research Laboratory, Human Performance Directorate, 711HPW/RH*

ABSTRACT
We describe a prototype cognitive work aid for airlift mission allocation and scheduling. A key design challenge was how to capture, represent, and utilize human-generated planning constraints that need to be respected when automatically replanning across multiple missions in response to situational changes. User feedback solicited via a formal user evaluation confirmed that the visualization and control mechanisms provided enabled them to understand and control the plans generated by the automated scheduler, and that the prototype aid allowed them to better assess and respond to large situational changes with across mission impacts. The work provides concrete illustrations of methods for externalising planning constraints so that they can be recognized and respected across distributed planning agents both human (airlift planners) and machine (automated scheduler). It serves to extend the range of available techniques for design of more effective joint cognitive systems.

KEYWORDS
Command and control; Transportation; Common Ground; Joint Cognitive Systems; Human-Automation Integration; Planning;

INTRODUCTION

This paper presents our most recent work leveraging automated support for airlift mission planning and scheduling. Our approach combines visualization and user-interaction principles to foster more effective joint performance between automated planners and schedulers and human users that together constitute the *joint cognitive system* (Woods & Hollnagel, 2006). Over the past decade we explored a variety of concrete techniques for design of joint cognitive systems across multiple decision-support development projects (DePass, Roth, Scott, Wampler, Truxler & Guin, 2011; Scott, Roth, Truxler, Ostwald & Wampler, 2009; Truxler, Roth, Scott, Smith and Wampler, 2012). These have included techniques to enhance *observability,* by visually representing the problem to be solved and candidate solutions in a form that users can easily understand, evaluate, and contribute to; and techniques to enhance *directability,* through control mechanisms that enable users to bound and direct the automated solution generation process (Roth, DePass, Scott, Truxler, Smith, and Wampler, in preparation). The goal is to enable the joint system to perform better than either the person or the automation could on its own.

The present application posed new design challenges that served to extend our cadre of techniques for observability and directability. In particular it required planning constraints that were implicit in the air mission plans generated by human planners to be exposed and explicitly represented so as to enable the automated planner to respect these constraints when situational changes dictated the need for dynamic replanning. In this paper we describe the design challenges, an initial prototype we developed and tested that begins to address these design challenges, as well as future directions.

BACKGROUND

Our work concerns the joint process of planning, scheduling, and execution of air missions by two cooperating military organizations. USTRANSCOM is charged with directing and executing the overall transportation needs for deployment of troops and distribution of goods via air, sea, and ground movements. The Air Mobility Command (AMC) is charged with executing the air movements. The first stage, referred to as *early planning,* begins at the point at which air movement requirements are locked in to the enterprise three weeks before movement. At that stage, planners at USTRANSCOM examine the entire known set of requirements, and match those against the limited resources they have to plan with – the number of aircraft that are scheduled to be available to AMC day by day, and the available throughput capacities of the airfields to be used to deliver those requirements. The second stage, detailed planning (also known as scheduling) is the focus of the current work. Once requirements are released by USTRANSCOM, detailed planners at AMC produce schedules of air missions to be flown, and maintain those schedules in an executable state (as changes happen) until the schedules are turned over to the AMC execution cell twenty-four hours before the air missions take off. The third stage, execution, covers the period from 24 hours before takeoff until the end of each air mission, and is handled by a group of Duty Officers (DOs) in AMC. In previous projects we developed systems to support both early planning (DePass et al, 2011 ; Truxler et al., 2012) as well as dynamic rescheduling during execution (Scott et al., 2009). The work described in this paper is aimed at improving detailed planning at AMC, both in terms of ease of producing the schedules and the quality of the schedules turned over to the AMC execution cell.

CURRENT PRACTICE

The scope of the decisions that must be made by AMC planners, both in terms of numbers of decisions and complexity of decisions is far greater than either of the other two stages. Further the quality of decisions made in this stage has the largest effect on the stated enterprise goals of effectiveness (delivering requirements on time) and efficiency (at a minimal cost). To transition from individual air movement requirements to executable mission schedules, there are four types of decisions to be made, that are split between two offices within AMC.

1. The first decision, known as *aggregation*, identifies a set of air movement requirements (or the pieces of a single air movement requirement) to occupy a single aircraft.
2. The aircraft resources available to planners are divided into wings, groups of like aircraft which share a common home airfield base. The second decision, known as *allocation*, determines the wing from which to source an aircraft for a particular aggregation of requirements as well as the allocation interval – the set of days for which an aircraft from that wing will be available for this use.
3. Air movement requirements specify cargo or passengers to be picked up at one location and dropped off at a second location. While not always possible, planners will try to link two or more air movement requirements together to be serviced by the same aircraft, in order to reduce the number of hours an aircraft will fly empty. The decision on how to combine movements together into a more efficient home-base to home-base route is referred to as *chaining*.
4. Finally, the creation of a detailed schedule for an entire home-base to home-base route is referred to as *scheduling*. Scheduling requires selection of en route stops to be made for refueling or crew rest, as well as determining takeoff and landing times for each of the flight legs (sorties). A host of details affecting timing must be taken into account in scheduling, including operating hours for each airfield to be visited, limited airfield parking capacity, and regulations on crew duty day lengths.
5.

The four decisions above cannot be made independently. Making the aggregation decision entails knowing what kind of aircraft is available, in order to know how much cargo can be put on the aircraft. The aggregation decision may need to be revisited if aircraft of that type turn out not to be available. Scheduling requires having aggregation, allocation, and chaining decisions already made – but difficulties in scheduling arising from overuse of enroute port capacity, for example, will require increasing the allocation interval for an aircraft, possibly leading to the aircraft not having suitable availability at all.

Making the decision process more complicated is the fact that the underlying planning state is not static. Both the numbers of aircraft scheduled to be available and the operating hours of the airfields can change with little notice. The air movement requirements that in theory were locked in three weeks before movement, in fact, are not static either. Requirements may be cancelled, they may change their schedule (as the availability of the cargo changes), they may change the amount of cargo to be carried. And, of course, as the enterprise reacts to military contingencies, to disaster relief and humanitarian assistance needs, late requirements (which tend to be high priority) can and will be added in to the mix at any stage. Adding further complication, these decisions are made by two independent, but cooperating, AMC offices. *Mission planners* are responsible for receiving the requirements from USTRANSCOM and creating the executable mission schedules that are eventually published to the AMC execution group. *Barrels,* who liaise between AMC and the wings decide which wing (and type of aircraft) a mission will be assigned.

In the current practice, the process followed is roughly linear and time consuming. First the aggregation decision is made by a mission planner. The mission planner may then ask a barrel planner for an allocation decision, or may wait a day or two while he looks for an appropriate second aggregation to pair it with (effectively making the chaining decision) before asking the barrel for an allocation.The barrel planner will make the allocation decision, and may suggest a chaining solution if the mission planner has not already done so. The mission planner will then prepare the detailed schedule, respecting aggregation, allocation, and chaining decisions.

While this process works, it leads to overall slates of mission schedules that are neither as efficient (cost-wise) nor as effective (in terms of delivering requirements on time) as they could be. These deficiencies stem from two factors. First is the sequential nature of the current process. When the barrel planner is asked for an allocation decision for a particular mission, he makes the best decision he can at that time, with the information he has available at that time. It might be that the very next allocation decision he is asked for would make better use of the aircraft he has just allocated to this mission, and an overall better slate of mission schedules would result if he changed his earlier allocation decision. Second, in the current process, decisions are unlikely to be revisited. Both mission planners and barrel planners are reluctant to remake a decision, for three reasons :

- Lack of time. Both mission planners and barrel planners are kept quite busy.
- Lack of visibility into how a change one planner would make would affect other planners. Either type of planner is reluctant to make a change that would cause potentially time-consuming rework for the other type of planner.
- Lack of cuing as to what the value of changing a decision might be. Even given their lack of time, if the planners could see what the value of a change would be, they would make time for the more valuable changes.

DESIGN OBJECTIVES AND CHALLENGES

The design objective was to leverage automated scheduling software to enable more rapid and efficient mission allocation and scheduling. The goal was to make more effective use of the limited airlift assets (e.g., reduce empty flying hours) and reduce overall costs (e.g., reduce overall flying hours which equates to reduced operating and fuel costs). More particularly, the objective was to facilitate replanning across multiple missions (mission reallocation and scheduling) when situational changes (e.g., an airfield closure ; a new, high priority emerging requirement) necessitated revisiting prior allocation and scheduling decisions to make most efficient use of assets, meet as many requirements as possible, and minimize overall costs.

To achieve these goals, we needed to address the factors we'd identified as obstacles to efficient replanning in current operation. We needed to design a system that allowed barrels and mission planners to more rapidly plan (and replan) missions than is possible with today's tools . The system also needed to make clear the changes made to missions, particularly changes that affected prior decisions made by others. Finally, the system needed to provide clear indication of the impact of the replan on high level efficiency (e.g., total flying hours ; empty flying hours) and effectiveness (e.g., number of missions delivering late) metrics. The system description below explains how all these design objectives were met. One of thc most significant technical challenges we faced was how to capture, represent, and utilize planning constraints that Barrels and Mission Planners consider in allocating assets and developing detailed schedules that were not explicitly represented in a system. Examples include cases where a mission was assigned to a particular wing because it required a plane type that was only available at that wing; and cases where a mission had to stop and refuel at a particular location because other refueling locations were temporarily dedicated to a different mission type (e.g., Ebola humanitarian aid missions). These detailed, mission-specific, planning constraints are not currently formally captured (appearing in comments sections if anywhere). The consequence is a lack of common ground with respect to relevant mission constraints across the distributed planning team (i.e., the Barrel, the Mission Planner, and the DO on the execution floor). The lack of common ground results in inefficiencies and scheduling errors, particularly when dynamic replanning, by someone other than the original planner, is required. The consequences of lack of common ground is amplified when one of the elements of the distributed planning team is an automated scheduler that is being relied upon to reschedule across multiple missions in response to situational events that create plan perturbations. In order to enable robust automated replanning that respects important (currently implicit) constraints it became necessary to develop mechanisms to allow Barrels and Planners to explicitly communicate planning constraints. These could be then be externally represented and used by the automated scheduler, effectively creating common ground across the broader distributed team that includes the automated scheduler, for more effective joint human automation planning.

In the next section we describe the prototype that was built and illustrate how it enabled users to communicate constraints that were then respected by the automated scheduler when a large-scale replan was required.

DESIGN FEATURES

We designed and implemented a prototype cognitive work aid to be used by both Mission Planners and Barrels. Each is supported with dynamic visualizations specifically designed for their needs. Mission planner visualizations display full details of individual mission schedules, and offer the ability to drag pieces of mission schedules with immediate feedback as to constraint violations. Barrel visualizations are organized around how the aircraft assets for particular wings are currently allocated, again offering Barrels direct graphical manipulation of elements of the plan with immediate alerting to conflicts. Our prototype offers a wide range of capabilities – we cannot discuss the full set of features in this paper. Here we concentrate on the design features that enable users to communicate planning constraints that are then explicitly represented in shared visualizations and respected by the automation when called on to replan.

In the simplest case a user can request the automation to schedule a new mission given a requirement (what to move, starting location, destination and by when it needs to arrive) and the scheduler will build a mission taking into account the current missions, wing allocations and airport constraints, possibly aggregating with other existing missions that match well in time and location. The resulting detailed mission can then be modified by the user; for example, a user may drag a sortie (a flight betwen two ports) in a timeline view to manually indicate when a sortie should arrive or depart. The system will adjust the other sorties accordingly making sure to immediately present any violations caused by the new schedule.

Importantly, the user can explicitly indicate a number of constraints that the scheduler needs to respect before asking the automation to schedule or reschedule one or more missions. Such constraints become directives and restrictions communicated to the scheduler which must be incorporated into the mission schedules being generated. For example, a constraint may specify that the mission must use an allocation from a certain wing. Another constraint is to specify that a sortie must arrive at a port within a certain time window (e.g., to account for customs and overflight clearances) ; or that it can only remain on the ground at the airbase for a limited amount of time. Other constraints include that a mission may not refuel at a certain location or must refuel at a certain location. Similarly they can specify a particular location for a rest stop or indicate that a particular location cannot serve as a rest stop.

In all of the above cases, it should be noted that the user can always manually adjust schedules output by the automation or even specify additional constraints and retask the automation to incorporate those changes into a revised mission schedule.

FUNCTIONAL EXAMPLE

The following series of screenshots illustrate how users would interact with the prototype to respond to a new emerging requirement that creates an over demand for available airlift assets requiring a broad across-mission reschedule. They illustrate how visualizations enable users to assess repercussions of situational changes, use the automated scheduler to revisit and repair previously scheduled missions, and direct which missions can be changed and how, by placing explicit constraints that are then respected by the automated scheduler.

Figure 1. Asset Dashboard

Visualizing impacts of emerging requirements on the ability to meet commitments. A Barrel has learned of some high priority missions about to be planned that will require 6 aircraft from the Charleston (KCHS) wing. To understand the intial implications, the barrel adds a reservation for 6 aircraft for the allocation window he believes will be needed to support the missions in the KCHS wing and looks at the initial impact. As shown in Figure 1, the prototype visualization will immediately show that this will overcommit tails at KCHS on days 335 – 338 (depicted as red boxes with negative numbers).

Assessing impacts on particular missions. As shown in Figure 2, the user can expand a given wing to view the currently allocated missions against the wing in a timeline form. From this timeline view he can observe which missions are impacted by the overcommitment (i.e., the missions with red dots under the

KCHS wing) and he can look at more detailed mission information via tooltips and other gestures on the timeline such as the cargo pickup locations for the missions, their allocation windows, their priority, and required delivery dates.

Revising mission schedules to resolve over commitment. The barrel could start to manually resolve the overcommitment by sliding mission times off to the right to avoid the overcommitment time, or rewing specific missions, noting violations that will appear as he modifies schedules. However, given so many missions are affected the Barrel may prefer to rely on the automated scheduler to rapidly generate a new across-mission schedule that minimizes mission delays.

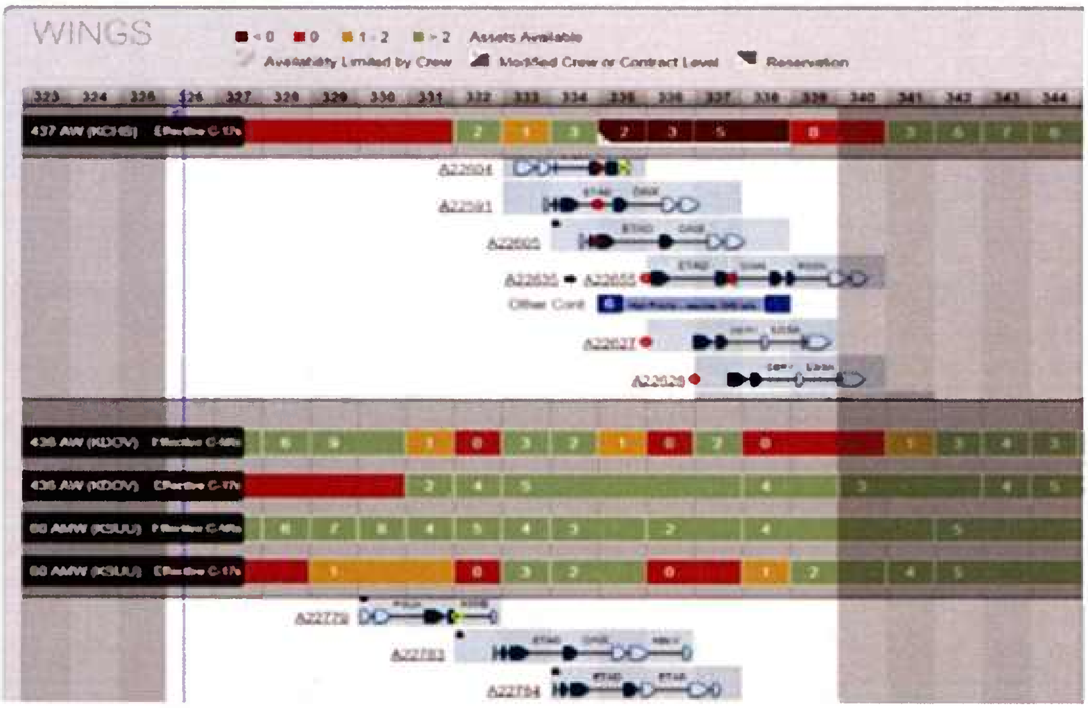


Figure 2.

Defining constraints to be respected by the automated scheduler. The barrel may choose to define constraints to be respected by the scheduler before calling it to resolve the overcommitment. For example, he can specify that some missions out of KCHS must remain where they were originally scheduled (locked in place) even if they coincide with the new high priority missions. Figure 2 provides some examples where the Barrel specified that missions be locked in place. These are indicated by the lock icon next to the mission schematic. As an alternative constraint, the Barrel can specify that a mission needs to continue to come from a particular wing (e.g., because only that wing has appropriately configured tails), but that it can slide in time.

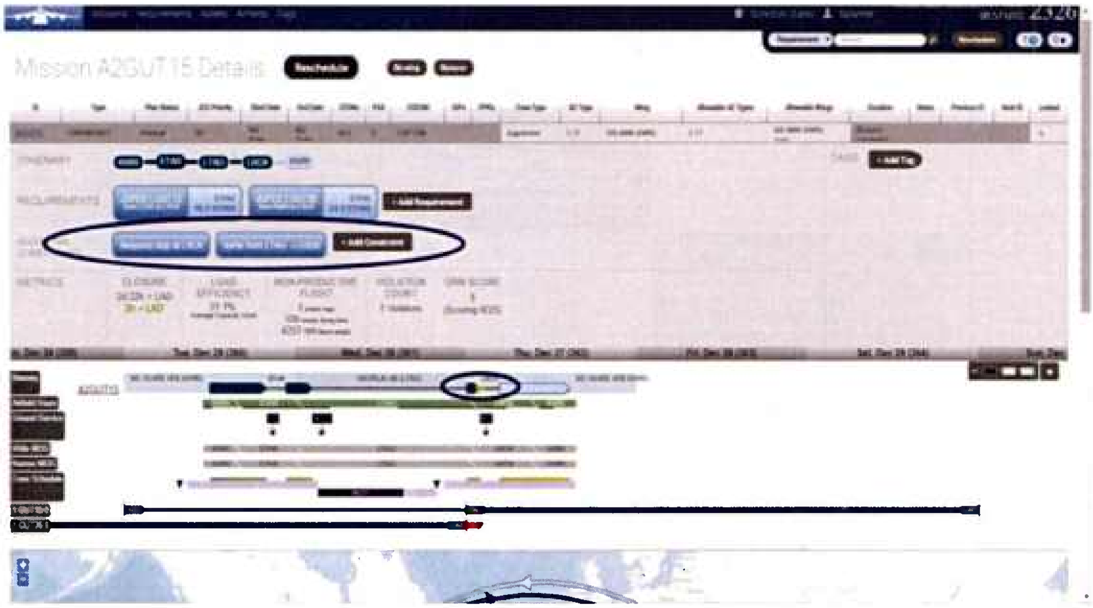


Figure 3: Detailed Mission Planning View with Mission Constraints

In addition to imposing constraints on wing allocation, the Barrel (or a Mission Planner) can impose constraints on individual mission schedules. Figure 3 shows a screenshot of a Detailed Mission Planning view for a single mission with multiple constraints : a timing constraint related to flight time of a specific sortie (LTAG to LRCK), and timing constraints related to a required mission stop (at LRCK). If the shifting of missions to accomodate the additional demand on aircraft out of KCHS starts to affect missions out of other East Coast wings such as McGuire (KWRI), this mission could be rescheduled, but the constraints entered by the original mission planner would be respected in the new schedule generated by the automated scheduler.

Inspecting and evaluating the revised plan produced by the scheduler. Once the user(s) have defined all constraints, the automation can be invoked to reschedule missions to accomodate the overcommitment of aircraft out of the Charleston wing. The automation will shift missions to other wings where possible, shift missions later if their required delivery dates allow, and sometimes shift missions such that they are late delivering their requirements, if unavoidable. As illustrated in Figure 4, the Change Summary views available after a reschedule will detail the types of changes made, their impact on overall metrics and the details of each mission change, allowing a planner or barrel to evaluate the solution as a whole or by specific mission changes. As shown in Figure 4, the Before Rescheduling column indicates the base state, that is, *before* the user asked the automation to reschedule. The right column provides details of what was changed when the automation rescheduled missions, taking into account the new reservation and various barrel wing and mission constraints. You can see there were 5 mission pairing changes – these indicate missions were chained together which will increase efficiency, there were 12 allocation interval changes – meaning the allocation duration for which aircraft will be reserved for missions from various wings changed. Looking further down, 9 missions had their wing assignments changed while others had sortie details such as refueling or rest location changes. 3 missions are now delivering past their required delivery dates.

Figure 4. Change summary view.

The Change Summary view also summarizes impact on overall efficiency metrics. Note that in this case the metrics at the bottom of the view which consider all missions in the system did not experience large perturbations as a result of the reschedule. In fact, empty flying miles and total miles were reduced likely due to the newly chained missions.

While the summary view provides a broad overview, planners whose missions are impacted would need to inspect and evaluate the mission changes in more detail. To delve into the details the user can click on either the category of interest or the Total Changed Missions row (12 Missions with Changes in the example) to see a multi-mission timeline view of each mission with their before and after schedules. The screenshot below shows details on five missions with changes – 1 row contains the details for each mission changed and if the mission was chained with another the timeline will represent each mission before chaining as well as the resulting chained mission. The left column provides a quick look indication of the categories of change that apply to that mission.

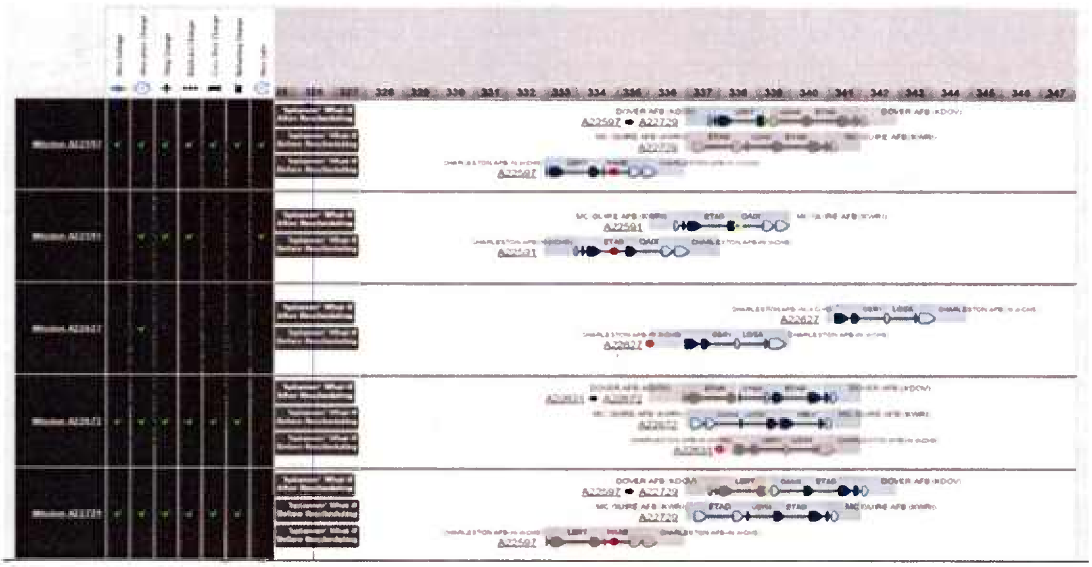


Figure 5 : Changed Mission Details

As the example illustrates the set of visualizations provides the distributed team - Barrels, Mission Planners, and the automated scheduler - with a shared representation of the missions to be scheduled, the constraints that need to be respected, and whether any constraints are violated by the present schedule, supporting common ground. Automated scheduler technology is leveraged to enable more efficient and effective schedules, while providing users control mechanisms to direct the automation via explict representation of constraints to be respected.

USER EVALUATION

A user evaluation was conducted at the completion of this first development cycle. Nine current practitioners, 3 Barrels and 6 detailed Mission Planners participated. A live demonstration of the prototype was presented using representative scenarios. Feedback was obtained via a written questionnaire that included 8-point scale rating questions eliciting feedback on the usability and usefulness of the capabilities demonstrated, and open-ended questions soliciting suggestions for additional capabilities to incorporate into future iterations.

Participant feedback, as reflected in both verbal comments and closed-form rating questionnaire scores was highly positive. As shown in Figure 6 participants indicated that they were able to understand and control the plans generated by the automated scheduler, and that the prototype aid would allow them to better assess and respond to large situational changes that had across mission impacts.

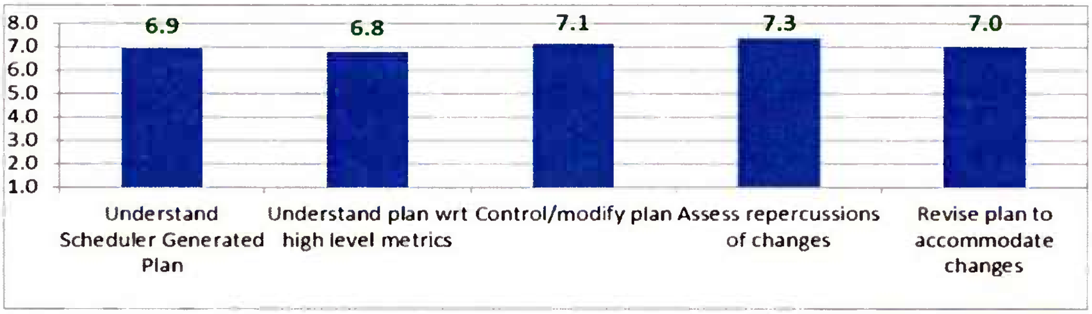


Figure 6. Mean rating on user evaluation questionnaire.

SUMMARY AND DIRECTIONS FOR FUTURE WORK

This paper provides an interim point description of our current application. Although our user evaluation results show that we have made significant progress towards delivering a mission scheduling capability to support AMC mission planners and barrel planners, we continue to improve system capabilities and particularly capabilities for capturing, representing, and utilizing mission planning constraints. To ensure the automated schedules meet the constraints known to the Barrels and Planners, it is imperative that Barrels and Planners be able to enter the constraints underlying the detailed planning decisions they consider to be important, sufficient so that any schedule that meets those constraints will be acceptable to the planners. This requires :

1. A sufficiently rich set of constraints for the planners to be able to detail their planning needs to the system.
2. A simple and effective mechanism to enter, visualize, and edit those constraints so that managing these constraints is not an undue burden on the planners.
3. The ability of the automated scheduler to respect these constraints in its production of schedules.

While we have made progress on this path, more research and development is required. Our future work will primarily focus on expanding the set of constraints available for planners to use to describe their planning needs.

The first broadening of the space of constraints will concentrate on adding conditions to the constraints. In the current system, for example, a planner might define a constraint that a particular airfield must be used as an enroute stop (for either refueling or crew rest). For the most part this unconditional constraint makes sense to the planners. But there are times when they would like to represent a more complicated pattern : Use airfield A as an enroute stop, as long as we are traveling east out of CONUS (or as long as we are using an aircraft from wings 1 or 2, or as long as we are starting this mission between days 52 and 54, to pick some more examples). And the planner might add an alternative – to use airfield B as the enroute stop if other conditions are met. The set of conditions available to planners clearly (given the examples above) will have to include methods of geographical reasoning, temporal reasoning, as well as simple boolean logic. The challenge in extending our system will be maintaining condition 2 – ensuring the ability of the planners to continue to manage this (expanded) set of constraints.

A second needed expansion of the constraint language arises from the fact that mission planners are not always planning individual air missions. They often schedule entire *movements* of air missions – some number of missions that are going to and from the same set of airfields, spread over some number of days. In the case of scheduling of larger movements, the mission planner can often identify particular constraints that apply to each of the individual missions that make up the movement. This leads to the desire for mission planners to be able to define a single constraint that will be applied to each of a set of missions. Our prototype system already has the notion of tags – one or more text strings attached to each mission. A set of missions, such as a movement, can be identified as the set of missions containing a particular tag. It will be relatively straightforward to allow constraints to be added at the tag level, instead of at the mission level.

While our projects have a primary aim of meeting our customers' needs, our objective is to also contribute to the generic corpus of reusable techniques for fostering more effective joint cognitive systems by making automated planners more observable and directable. In this case we provided a concrete illustration of methods for externalising planning constraints, so they can be recognized and respected across distributed planning agents both human (Airlift Planners) and machine (automated scheduler). It is our hope that the present work serves to extend the range of available techniques for design of collaborative automation.

REFERENCES

DePass, B., Roth, E. M., Scott, R., Wampler, J. L., Truxler, R., and Guin, C. (2011). Designing for collaborative automation: A course of action exploration tool for transportation planning. In *Proceedings of the 10th International Conference on Naturalistic Decision Making*, May 31-June 3, 2011, Orlando, FL. (pp. 95 -100)

Scott, R., Roth, E. M., Truxler, R., Ostwald, J., Wampler, J. (2009) Techniques for effective collaborative automation for air mission replanning. In *Proceedings of the Human Factors and Ergonomics Society 53rd Annual Meeting*. (pp. 202-206). Santa Monica, CA: HFES.

Truxler, R., Roth, E., Scott, R., Smith, S., and Wampler, J. (2012) Designing collaborative automated planners for agile adaptation to dynamic change. *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting (pp.* 223- 227) Santa Monica, CA: HFES.

Woods, D. D. and Hollnagel, E. (2006). *Joint Cognitive Systems: Patterns in Systems Engineering*. Boca Raton, FL: Taylor & Francis.

Developing a semi-automated HTA process

Neville A STANTON[a], Katherine L PLANT[a], Loukas RENTZOS[b], Charalampos VOURTSIS[b], Stratos ANTONIOU[b] and Konstantinos SMPAROUNIS[b]

[a] *Transportation Research Group, Faculty of Engineering and Environment, Boldrewood Campus, University of Southampton, Southampton, SO16 7QF, UK*

[b] *Laboratory for Manufacturing Systems and Automation, Department of Mechanical Engineering and Aeronautics University of Patras, Patras 26500, GREECE*

ABSTRACT

This early stage research describes the current efforts undertaken to semi-automate the Hierarchical Task Analysis (HTA) method as part of a European Union funded aviation project. HTA is one of the most popular and widely used Human Factors methods, however it can be laborious and time consuming to conduct, particularly in complex socio-technical systems such as the aviation environment. This early stage research paper and associated poster will describe the work undertaken to develop a semi-automated HTA procedure and will also consider options for capturing the cognitive elements of task analysis (e.g. decision making).

KEYWORDS

Decision Making, Methods, Manufacturing, Transportation,

INTRODUCTION

HTA has been described as one of the most popular and widely used Human Factors methods, partly owing to its flexibility and scope for further analysis that it offers (Stanton et al., 2013). HTA was developed in the 1960s to understand the skills required in complex, non-repetitive operator tasks. At the time, HTA was a radically new approach, based on functional, rather than behavioural or psychometric constructs (Annett, 2004a). The basic premise behind HTA is that tasks are explored through a hierarchy of goals indicating what a person is expected to do and plans indicate the conditions when subordinate goals should be carried out, which is akin to the decision making process of an operator. Each goal and the means of achieving them are represented as an operation. Stanton (2006) argued that HTA has three governing principles, which have remained unchanged in nearly half a century that HTA has been used. Firstly, HTA is proposed as a means for describing a system in terms of its goals, as such HTA provides a goal based analysis of a system and the system analysis is presented via the HTA. Secondly, HTA offers a means of breaking down sub-operations into a hierarchy and thirdly, rules (called plans) exist that the guide the sequence that the sub-goals are attained.

Due to its flexible nature HTA has seen many applications across a variety of domains, including interface design and evaluation, manual design, job description, training, allocation of function, job aid design, error prediction and analysis, team task analysis, workload assessment and procedure design (Stanton, 2006). However, one of the biggest limitations of the method is that it can be laborious and time consuming to conduct. This is particularly evident in complex socio-technical systems such as the aviation domain, where the naturalistic decision making (NDM) environment is epitomised with tasks being conducted in dynamic conditions often with limited time and goal conflicts. These attributes of an NDM environment can result in large and unwieldy HTA outputs even for relatively simple tasks. Previous research by the Human Factors Integration-Defence Technology Centre sought to address this with the development of a HTA software tool. The tool ran on the Microsoft.net framework and used a familiar windows-based interface which interfaces directly with Office applications such as Word. The software tool provided structure and expedited the documentation and presentation of the analysis results, allowing for quick edits to be made that propagated through to the rest of the analysis. However, this tool is still a desktop version of the pen and paper method and therefore requires a high proportion of manual input from a Human Factors analyst. As part of the EU funded i-VISION (Immersive Semantics-based Virtual Environments for the Design and Validation of Human-centred Aircraft Cockpits) project work has been undertaken to automate the HTA process in a virtual reality (VR) environment.

Despite the widespread application of HTA it is often assumed that HTA is unsuited for dealing with cognitive tasks (Shepherd, 1998). Generally, HTA is seen as way to establish an accurate description of the steps required to complete a task whereas the focus on cognitive task analysis (CTA) is to capture the representation of knowledge that people have, or need to have, in order to complete tasks. Shepherd (1998) argued that despite this distinction, successful performance in all tasks depends upon the interaction between physical and cognitive elements. As such, Shepherd (1998) advised that rather than distinguishing between cognitive and non-cognitive task analysis, one should consider how a general task analysis strategy accommodates cognitive tasks. HTA can account for cognitive aspects of task performance in two ways. Firstly, HTA can be adapted to incorporate elements from CTA methods and secondly, cognitive task elements can be inferred through HTA by stating plans as these enable the decision process to be inferred even if decision making was not apparent through observation or discussions with operators (Shepherd, 1998). A secondary aim of the i-VISION project work is to determine how cognitive tasks (e.g. decisions) can best be represented in a virtual reality environment. Options for this will be discussed.

METHOD

As an initial case study a procedure from the manufacturing industry was selected (differential gear assembly in automotive manufacturing), subsequent research efforts will develop this work within the context of the aviation domain. The case study selected for this initial development was that of the assembly of a critical component (gear differential) in car production. This task was selected due to its simplicity and because it offers the opportunity of multiple abstraction levels once the basic HTA is done. The differential consists the casing and the following parts: the pinion, the engine axle mount, the drive axle mount, the drive axle mount flange, the crownwheel, the differential cap, and bolts and screws. The parts are placed in the case one after another with one exception where one part is placed on another and then the system of two parts is attached in the case.

HTA development

A manual HTA was constructed for the task under analysis by two Human Factors experts and three systems manufacturing experts. This enabled the identification of verbs to describe the physical tasks that were performed in this assembly of the differential gear. Rules based on the VR principle of collision detection were written for each verb and implemented into the VR system, along with tagging objects, tools used and changes of object states. Verb definitions were determined by the Human Factors and Systems Manufacturing domain experts and task time literature was utilised where relevant. Table 1 provides some examples of the physical verbs that were defined, how these would be detected in the VR environment and the resulting object states.

Table 1. Examples of physical verbs for the automated HTA process

Verb	VR detection	Object state
Touch	translation [hand] + collision detection [hand + object or surface]	contact with object or surface, object state doesn't change, i.e. no movement
Press	TOUCH + continuous pressure >1000ms + RELEASE	change of position >1000ms
Hold	TOUCH + closing fingers around object >1000ms	change of position [fingers], close around [object] >1000ms
Move	translation [any body part]	change of positon [body part]
Assemble	TOUCH + GRASP + MOVE + collision detection [3 or more objects]	change of position [objects], contact between 3+ [objects]
Insert	TOUCH + GRASP + MOVE + collision detection [object + object] + PRESS	change of position [objects], internal contact between [object] + [object]

Technical description

In this work, HTA has been modelled in order to work as an integrated part of a VR environment. A virtual platform was used in order to program the HTA. The VR method produced uses an algorithm developed to generate each task based on the human user's motion and their interactions (i.e. collisions) with several elements of the virtual product (Rentzos et al., 2014). The extraction and storage of the HTA is accomplished by using and manipulating arrays inside the VR platform. The main virtual environment interaction principle that is used in this development is collision detection. Collision detection identifies whether or not two or more virtual elements are 'colliding' each other. Additionally, a principle called "magnets" was used in order to simplify the virtual task for the user. The working principle of the magnets method lies in identifying the proximity of an object with another in order to position them. When an object approaches the one closest to its final position, the approaching object automatically is positioned in its final position.

The main aim was to automatically generate the HTA verbs (Table 1) in the virtual environment, in order to have a tool that would produce a valid HTA. During the development and testing, only physical interaction monitoring was used. For example, when a user grasped an object, it was assumed that the object had previously been identified among the other objects. It is intended that other modalities (including visual and auditory) will be included in future iterations with the integration of relevant technologies such as eye tracking and audio devices. Therefore, the 'identify' action will be defined through eye tracking, rather than making the assumption that this has occurred. Aside from simple verb extraction from physical collisions in the virtual environment, many verbs were recognized by identifying whether two or more extracted verbs where performed simultaneously in the VR environment. For example, when Touch, Grasp and Move were performed simultaneously, it is assumed that the user was performing the verb Carry.

The algorithm begins by monitoring the assembly process. Every time the users hand (which is defined an object) collides with a virtual object a verb-task is generated which corresponds to the action performed by the user. This way a readable sentence with correct grammar is created for each user interaction (corresponding to the task step terminology that would be recorded by the HTA expert in the manual method if the task had been observed). By utilizing all of the above information in combination with the Hierarchy Manager of the VR platform we are able to clarify levels of abstraction for each task. This way a complete HTA tree was automatically extracted without any intervention from the VR expert, although they can monitor the process and correct faults that the machine cannot detect. Figure 1 shows the disassembled components of the differential (as they appear in the virtual environment) and the generated HTA is shown on the right.

Figure 1. Population of the hierarchy array

CURRENT RESULTS AND FUTURE WORK

Semi-Automating the HTA process: Manufacturing case study

The HTA is automatically populated by a user performing tasks in the VR environment. Visual instructions are generated after the user has performed a task (Makris et al., 2012). The programmed HTA was capable of representing tasks at different levels of abstraction. For example, most of the tasks involved placing a part (e.g. pinion) into the casing one after another. However, there is the instance where one part is placed onto another and then this newly assembled part is put into the casing (steps 3.1.1, 3.1.2 and 3.1.3 in Figure 2). At this point the tasks are moved one level deeper in the hierarchy. Figure 2 provides the completed HTA output from the automated procedure (as an array of the VR platform). From this array, formal sentences can be extracted and the results can be saved in an editable file e.g. xml. The ambition is for this method in other domains, such as aviation, to assist with the rapid design and prototyping of novel concepts.

1. HTA ID	2. Verb	3. Object 1	4. Object 2	5. Tool
1	ATTACH	pinion	Case	Right Hand
1.1	Move	pinion	Parts Table	
1.2	Grasp	pinion		Right Hand
1.3	Move	pinion	Assembly Table	
1.4	Attach	pinion	Case	Right Hand
2	ATTACH	Engine axle mount	Case	Right Hand
2.1	Move	Engine axle mount	Parts Table	
2.2	Grasp	Engine axle mount		Right Hand
2.3	Move	Engine axle mount	Assembly Table	
2.4	Attach	Engine axle mount	Case	Right Hand
3	ATTACH	crownwheel	Case	Right Hand
3.1	ATTACH	flange	crownwheel	Right Hand
3.1.1	Move	flange	Parts Table	
3.1.2	Grasp	flange		Right Hand
3.1.3	Attach	flange	crownwheel	Right Hand
3.2	Move	crownwheel	Parts Table	
3.3	Grasp	crownwheel		Right Hand
3.4	Move	crownwheel	Assembly Table	
3.5	Attach	crownwheel	Case	Right Hand
4	ATTACH	Drive axle mount	Case	Right Hand
4.1	Move	Drive axle mount	Parts Table	
4.2	Grasp	Drive axle mount		Right Hand
4.3	Move	Drive axle mount	Assembly Table	
4.4	Attach	Drive axle mount	Case	Right Hand
5	ATTACH	cap	Case	Right Hand
5.1	Move	cap	Parts Table	
5.2	Grasp	cap		Right Hand
5.3	Move	cap	Assembly Table	
5.4	Attach	cap	Case	Right Hand

Figure 2. Output from the automated HTA procedure

Representing cognitive tasks in the HTA process

We have been able to semi-automate the HTA process for physical tasks. The next challenge is representing other modalities. It is envisaged that visual, auditory, vocal tasks will be integrated with the introduction of other technologies including eye tracking and voice recognition. However, finding the best way to represent cognitive tasks is still under consideration and has been a long standing challenge for task analysts (Phipps et al., 2011). Task verbs in the cognitive modality include: select, store, recall, decision, recognise, and identify. A study by Phipps et al. (2011) evaluated two extensions to HTA: sub-goal templates and the skills-rules-knowledge framework, for analysising cognitive activity in anaesthetic tasks. They found that both provided qualitatve insights about cognitive performance. However, the extended methods involved extensive manual classification which negates any advantages in terms of time and effort saved in the production of the automated HTA for physical tasks.

The introduction of cognitive probes at various points during task completion is potentially a less resource intense method of gathering cognitive information. Selected probes could be introduced to the simulated task activity either in a freeze-probe format via pop-up questions in the VR environment. Schutte and Trujillo (1996) integrated probes into flight deck scenarios via naturally occuring conversations with airtraffic controllers and dispatchers. Alternatively, cognitive tasks could be assessed via an interview with the operator at the end of the actvity and the automated HTA could be manually edited by the analysts. This approach would still save time over the traditional HTA method.

CONCLUSION

The early-stage research presented descrives the efforts that have been undertaken to automate the HTA process. Initial implementation in a manufacturing case study has been successful and we move towards applying a similar process in the aviaition environment. Our next stept is to dertmine how cognitive tasks are best captured and this will commence with exploring the implementaion of cognitve probes via naturally occuring conversations on the flight deck as an initial means of gathering cognitive information.

STATEMENT OF INNOVATION

To our knowledge, this is the first time that semi-automating the HTA process has been documented in the literature. This is currently early stage research but initial results have been encouraging and will expand into the aviation context. We envisage that the final project outputs will have far reaching applications by modernising and expediting the HTA process and therefore will have the potential to benefit many Human Factors researchers and practitioners.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 605550.

REFERENCES

Annett, J. 2004a. Hierarchical Task Analysis. In D, Diaper & N.A. Stanton (eds.) *The Handbook of Task Analysis for Human-Computer Interaction.* Mahwah, NJ: Lawrence Erlbaum.

Phipps, D.L., Meakin, G.H. & Beatty, P.C.W. 2011. Extending hierarchical task analysis to identify cognitive demands and information design requirements. *Applied Ergonomics*, 42, 741-748.

Makris, S., Rentzos, L., Pintzos, G., Mavrikios, D., Chryssolouris, G. 2012. Semantic-based taxonomy for immersive product design using VR techniques. *CIRP Annals - Manufacturing Technology, Volume 61(1), 147–150*

Rentzos, L. Vourtsis, C., Mavrikios,D., Chryssolouris,G. 2014. Using VR for Complex Product Design , Virtual, Augmented and Mixed Reality. *Applications of Virtual and Augmented Reality, Lecture Notes in Computer Science, Volume 8526,455-464 (2014)*

Schutte, P.C. & Trujillo, A.C. 1996. Flight crew task management in non-normal situations. *Proceedings of the 40th Annual Meeting of the Human Factors and Ergonomics Society*, pp. 244-248. Santa Monica, CA: HFES.

Shepherd, A. 1998. HTA as a framework for task analysis. *Ergonomics*, 41(11), 1537-1552.

Stanton, N.A. 2006. Hierarchical Task Analysis: Developments, applications and extensions. *Applied Ergonomics*, 37, 55-79

Stanton, N.A., Salmon, P.M., Rafferty, L.A., Walker, G.H., Baber, C. & Jenkins, D.P. 2013. *Human Factors Methods. A Practical Guide for Engineering and Design.* 2nd ed. Aldershot: Ashgate.

The Influence of Operating Room Handoffs on Teamwork, Stress, and Work: a 360 degree Evaluation of Team Shared Situation Awareness

Cristan E. ANDERSON MD [a,b] and Lygia STEWART MD [a,b]

[a]*Department of Surgery, University of California, San Francisco*
513 Parnassus Avenue, Room S-321, San Francisco, CA 94143

[b]*San Francisco VA Medical Center*
Surgery (112), 4150 Clement Street, San Francisco, CA 94121

ABSTRACT

Teamwork is important in the operating room. Team members rely on each other's expertise for successful task completion. Because of lunch and change of shift breaks, handoffs occur frequently, but little is known about their effect on team performance, or the team's awareness of these changes. We performed a 360 degree evaluation of the effect of operating room handoffs on teamwork, stress, and work among the members of the operative team (surgeons, anesthesia providers, circulator nurses, scrub technicians). An independent observer also evaluated the effect of handoffs. We specifically examined for evidence of shared team situation awareness. Surgical attendings reported decreased teamwork, increased stress, and increased work due to handoffs in about 30-50% of cases; while nursing personnel reported handoffs to be seamless, and have little effect (5%) on teamwork, stress, or work. This demonstrated a lack of shared team situation awareness, among operating room teams, regarding the influence of handoffs on team performance.

KEYWORDS

Surgical teams, Handoffs, Situation Awareness, Patient Safety

INTRODUCTION

The operating room (OR) is a complex environment in which effective communication and the coordination of multiple team members is crucial for safe and efficient functioning. Team members rely on one another's expertise for completing tasks successfully. They must share information rapidly when responding to expected and unexpected events. Many have suggested that aviation and surgery share common features, and suggested using aviation safety procedures to provide a framework for quality improvement. (Hugh 2002; Karl 2009; Sexton, Thomas, & Helmreich 2000). And, medical team training (MTT) programs, developed utilizing concepts from aviation crew-resource management, including checklists and briefings; have led to improvements in team communication, team performance, decreased delays, and improved patient safety. (Wolf, Way & Stewart 2010; Neily, Mills, Young-Xu,, et al., 2010; Young-Xu, Neily, Mills, et al., 2011). Our group has had a robust MTT program, with >95% participation, present for over 8 years.

There are important differences between surgical teams and aviation teams. Aviation teams remain fixed (no other choice at 30,000 feet), and have a mandatory standardized experience on airplanes they crew. Surgical teams are composed of surgeons, anesthesia providers, a circulating nurse (who is outside the sterile field), and a scrub technician (who is inside the sterile field and works directly with the surgeons). Unlike aviation teams, surgical teams change during the case. Labor and union policies dictate two 15 minute breaks, and one 30 minute lunch break, in an 8 hour working period. Also, depending on staffing, emergencies, absences, etc., circulating nurses and scrub technicians may not have extensive experience (or in some cases, very little experience) with the specific case type they staff. While specialized nursing and anesthesia teams are commonly designated for certain surgical specialties (e.g., cardiac and transplant), this is not the case across the board. There is a general concept of cross-training in nursing; and it is common to work with non-specialized nursing staff, especially later in the day (after change of shift). Studies have shown that working with a fixed, specifically trained, nursing staff results in improved patient outcomes, improved safety climate, improved efficiency, and lower costs (Kenyon. Lenker, Bax, & Swanstrom 1997; Muller, Zalunardo, Hubner, Clavien, & Demartines 2009; Stepaniak, Heij, Buise, Mannaerts, Smulders, & Nienhuijs, 2012). But, even fixed teams require mandatory breaks (as dictated by labor law agreements).

As defined by Endsley (1995), shared situation awareness is the degree to which the team has reached a common state of understanding. There are no studies that have examined the effects of handoffs on team dynamics, or shared team situation awareness during surgical operative cases. We studied this..

METHODS

We performed a 360 degree evaluation of handoffs during surgical cases. Detailed evaluation questionnaires were given to all members of the OR team, including: Surgical attendings, Surgical residents, Anesthesiology providers (MDs, nurse anaesthetists), Circulating nurses, and Scrub nurses or technicians. Evaluations were directed toward a specific operative case, and they were completed in real time, immediately following the provider's portion of the case. All medical professionals who participated in the case were given evaluations forms, including surgeons; nursing and anesthesia team members who started the case, provided relief (breaks, lunch), or who completed the case following change of shift. In addition, 20 cases were observed and evaluated by an observer who did not participate in the operative case; this person was a surgeon with a focus in human factors.

Evaluations covered teamwork, specifics of the case, the providers role, stress from handoff, extra work related to handoffs, and the overall process. Evaluation forms were designed so we could correlate responses from the various medical professionals on the same case.

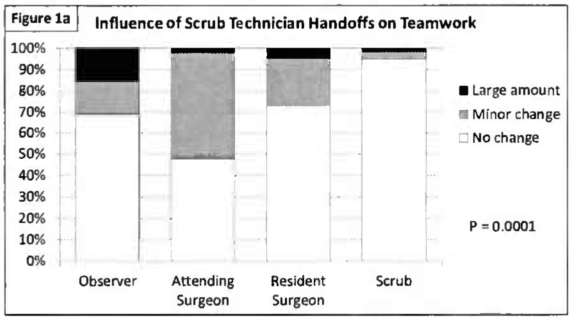
RESULTS

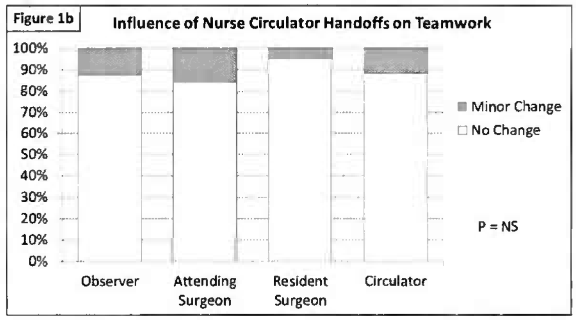
A total of 359 evaluation forms were completed, covering 89 different operative cases. The number of evaluations completed by each type of medical professional is shown in Table 1. The average number of handoffs / operative case by each medical professional group was as follows: Surgery 0, Anesthesia 1.2 (range 0-4), Circulator Nurse 1.7 (range 0-4), Scrub Technician 1.7 (range 0-5). This means that, during an average case, surgeons worked with a total of 2 Anesthesia providers, 3 circulator nurses, and 3 scrub technicians. There were no nursing handoffs in 12-15% of cases, and no anesthesia handoffs reported in 35% of cases; the maximum number of handoffs reported was: 5 Anesthesia providers, 5 circulator nurses, and 5 scrub technicians during one surgical case. Surgeons reported that the timing of circulator and scrub handoffs were optimal in only 69% and 55% of cases, respectively.

Table 1. Medical Professionals Participating in Handoff Evaluations

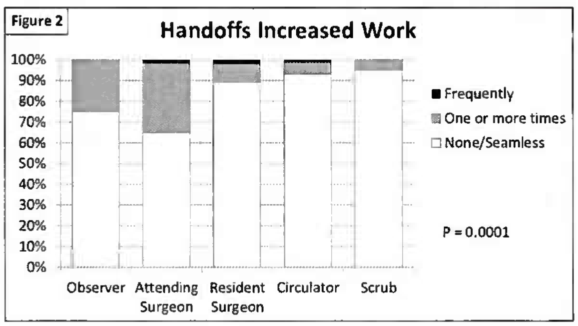
Medical Professional	Subgroup	Number
Surgeon	Attending	63
	Resident	46
	Observer	20
Anesthesia	MD / Nurse Anaesthetist	48
Nursing	Circulator	108
	Scrub Technician	74

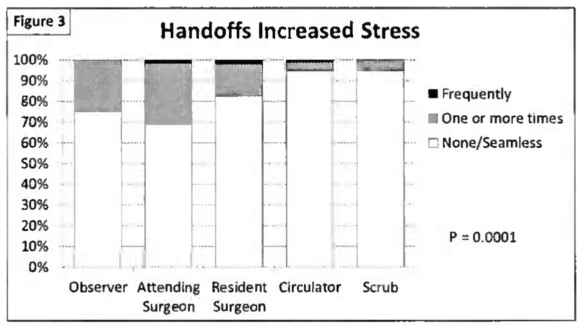
Anesthesia handoffs were generally reported to be seamless (96%) and were often (63%) not noticed by surgeons, including the observer. Analysis of Anesthesia handoffs were not part of the nursing evaluation. Anesthesiology providers evaluated their handoffs as 100% seamless, with no change in teamwork, stress, or work. Anesthesia, in all these cases, had a more separate defined role. Given these findings, we did not analyse Anesthesia handoffs in the context of the rest of the surgical team.



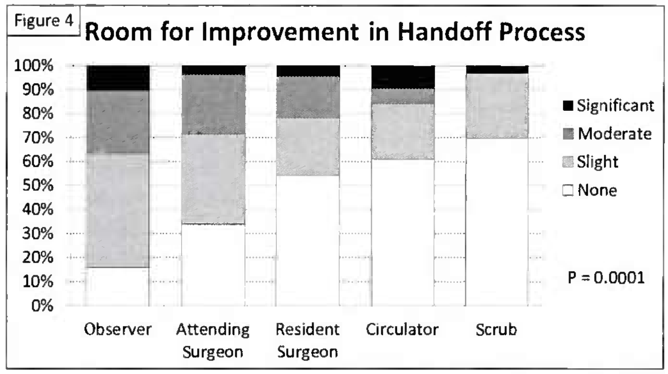


The influence of handoffs on teamwork is shown in Figure 1a and 1b, as reported by the various team members. As shown, attending surgeons reported that scrub technician handoffs changed teamwork more than 50% of the time, while the observer reported a 30% change; in contrast, scrub technicians perceived minimal changes in teamwork (5%); these differences were significant ($P < 0.0001$, Chi-squared). Nurse circulator handoffs, were reported to have a lesser effect on teamwork, and there were no significant differences in perception among the team members





The influence of handoffs on work and stress are shown in Figures 2 and 3, as reported by the various team members. As shown, attending surgeons reported that handoffs increased work, and created stress, in over 30% of cases. The observer was somewhat aware of the increased work and stress due to handoffs, but less than the attending surgcons. In contrast, nurse circulators and scrub technicians generally lacked awareness of the increased stress and work caused by handoffs; differences in awareness were highly significant ($P < 0.0001$, Chi-squared).



Evaluation of the overall handoff process by the various team members, regarding room for improvement, is shown in Figure 4. As can be seen, the observer perceived the largest need for improvement in the handoff process (85%), followed by the surgical attending (67%). Once again, nurse circulators and scrub technicians perceived the handoff process to be more flawless than other team members; the differences in awareness were highly significant (P<0.0001).

DISCUSSION

There has been considerable interest in physician handoffs during normal patient care, including the increased number of resident handoffs due to current limitations in resident work hours. These studies have highlighted possibile patient safety issues related handoffs. (Borman, Jones, & Shea, 2012; Lee, Myers, Rehmani, et al., 2012; Kitch, Cooper, Zapol, et al, 2008; Charap 2004). But, there are no studies examining the effects of handoffs in the operating room on patient safety. One reason this is an important issue is that handoff frequency is actually much higher in the operating room than with normal patient care on hospital wards. Most hospital patient-related handoffs occur at the end of an 8-12 hour physician shift; while in the operating room, cases lasting a few hours can be associated with multiple nursing personal handoffs, especially if they occur near the lunch hour or change of shift (usually 3 pm); and only 12-15% of cases, reported no operative nursing staff handoffs. Thus, handoff "density" is much higher in the operating room.

A number of studies have reported that working with designated, or specifically trained, operative nursing staff results in improved patient outcomes, improved safety climate, efficiency, and lower costs (Kenyon. Lenker, Bax, & Swanstrom 1997; Muller, Zalunardo, Hubner, Clavien, & Demartines 2009; Brown, Parker, Quiñonez, et. al., 2011; Stepaniak, Heij, Buise, et al, 2012). Also, an observational study of cardiothoracic surgery cases, reported that teams whose members were familiar with the operating surgeon had significantly fewer events and teamwork failures than teams where the majority of members were unfamiliar with the operating surgeon. But only 4% of failures were attributed to handoffs. (ElBardissi, Wiegmann, Henrickson, Wadhera, & Sundt, 2008). These studies highlight the advantages of stable specialized teams. But, many of these studies were in populations where specialized nursing staff (circulator and scrub technicians) are standard (eg, cardiothoracic surgery), this is not the case for many other surgical specialities. But even specialized teams have handoffs for lunch and change of shift breaks.

This study examined the influence of handoffs on teamwork. Teamwork is very important in the surgical arena (Weaver, S.J., Rosen, M.A., Diaz-Granados, D., Lazzara, E.H., Lyons, R., Salas, E., et al, 2010). This study was specifically designed to capture the perceptions of all the members of the surgical team. We noted that OR nursing personal were generally unaware of the deleterious effects of handoffs on surgeons, especially the attending surgeon. Attending surgeons reported that handoffs decreased teamwork, increased work, increased stress, were not at an optimal time in 31-45% of cases, and had significant room for improvement. Similar observations were made by the independent observer. In contrast, nursing staff reported that handoffs were usually seamless, did not decrease teamwork, and did not increase work or stress. Differences were highly significant. This demonstrates a lack of shared team situation awareness. It is perhaps an even more significant finding since this lack of awareness occurred in the setting of a medical center with robust medical team training protocols.

Shared situation awareness is defined as the overlap in individual situation awareness across a population at a given moment in time and space. High level shared situation awareness occurs when this overlap is robust, accurate, ongoing, and includes the necessary information for each individual person to perform his/her part in the overall group effort. This is crucial for success for small groups engaged in complex tasks. This is the case in surgery. The greatest loss in teamwork due to handoffs was between the surgeon and scrub technician (who works directly with the surgeon). Thus, following the handoff, the surgeon was now working with a team member who was less able to anticipate what is needed. This (anticipation) is an important non-technical skill for scrub nurses (Mitchell, Flin, Yule, et al, 2011). The surgical attending has the responsibility for the patient, while to some degree anesthesia and nursing personal are interchangeable (and they did change, as this study documented), the attending surgeon is not interchangeable. It was clear that the attending surgeon, who carries the responsibility for the patient, more acutely felt the impact of issues related to changes in OR personal; even more than other surgeons (surgical residents, independent observer). The surgical attending, working at the sharp end of the point, carries the brunt of the system issues.

An increase in stress during operative cases, even 30%, is meaningful. A number of studies have examined the influence of stress on individual and team performance. A large body of research indicates that the individual's breadth of attention narrows, they tend to become more self-focused. Group members adopt a narrower, more individual perspective of task activity, and with this narrowing of perspective, team members' cognitions shift from a broader, team perspective to a more narrow individualistic focus. (Driskell & Salas 1991, 1999). Additionally, stress has been reported to decrease surgeons non-technical skills, including communication and decision-making (Arora, Sevdalis, Nestel et al 2010; Arora, Hull, Sevdalis, et al, 2010). A number of studies have highlighted the importance of non-technical skills for surgeons (Youngson & Flin 2010; Flin, Yule, Paterson-Brown, et al, 2007; Flin, Yule, Paterson-Brown, et al, 2006).

It is important to put this study into the context of other studies employing a survey to determine physician attitudes about patient safety. This study was not a general survey about safety culture attitudes; instead we asked medical providers to evaluate a specific operative case, and describe the effect of the handoff *on that case*, and on the various aspects of team performance. Data was collected in real time, about specific operative cases. Interestingly, studies reporting physician safety culture attitudes reported that surgeons were more optimistic than nurses. These studies described substantial discrepancies in perceptions of teamwork held by surgeons and nurses; surgeons rated the teamwork of others as good, while nurses perceived teamwork as poor. (Makary, Holzmueller, Sexton, et al, 2006; Makary, Holzmueller, Thompson, et al, 2006; Sexton, Thomas, & Helmreich, 2000). The tendency for surgeons to rate teamwork positively, makes the findings in the current study even more meaningful.

CONCLUSION

Teamwork is important in the operating room. Team members rely on the specific expertise of other team members for successful operations. This study examined the effect of handoffs on team performance, and looked for evidence of shared team situation awareness of any handoff-induced variations in performance. A 360 degree evaluation of the effect of operating room handoffs on teamwork, stress, and work among the members of the operative team (surgeons, anesthesia providers, circulator nurses, scrub technicians) as well as evaluation by an independent observer was performed. Surgical attendings reported decreased teamwork, increased stress, and increased work due to handoffs in about 30-50% of cases; while nursing personnel reported handoffs to be seamless, and have little effect (5%) on teamwork, stress, or work. This demonstrated a lack of shared team situation awareness, among operating room teams, regarding the influence of handoffs on team performance. This is an important observation, since, in the setting of increased stress and work, surgeons need additional support from their team members. If the team members have no perception of the issues created by handoffs, they will not act to lend the support that is needed. This is the first, and only, study on the effect of handoffs on team performance in the operating room. The findings are significant and warrant an active response. It is hoped that dissemination of this information will increase awareness of this issue, and create opportunities to decrease this effect.

ACKNOWLEDGMENTS

We would like to thank all the surgeons, nurses, scrub technicians, and anesthesiology personnel who participated in our evaluations.

REFERENCES

Arora, S., Sevdalis, N., Nestel, D., et al (2010). The impact of stress on surgical performance: a systematic review of the literature. *Surgery*, 147, 318–330.

Arora, S., Hull, L., Sevdalis, N., et al, (2010). Factors compromising safety in surgery: stressful events in the operating room. Amer J Surgery, 199, 60–65.

Borman, K.R., Jones, A.T., & Shea, J.A., (2012) Duty Hours, Quality of Care, and Patient Safety: General Surgery Resident Perceptions. *Journal of the American College of Surgeons*, 215, 70-79.

Brown, M L, Parker, S E, Quiñonez, L G, et al. (2011). Can the impact of change of surgical teams in cardiovascular surgery be measured by operative mortality or morbidity? A propensity adjusted cohort comparison. *Annals of Surgery*, 253, 385-92.

Charap, M., (2004) Reducing resident work hours: unproven assumptions and unforeseen outcomes. *Annals of Internal Medicine*. 140, 814–815

Driskell, J.E. & Salas, E., (1991) Group decision making under stress. *J Appl Psychol*, 76, 473–478.

Driskell, J.E. & Salas, E., (1999). Does stress lead to a loss of team perspective? *Group Dyn*, 4, 291–302.

ElBardissi, A.W., Wiegmann, D.A., Henrickson, S., Wadhera, R., Sundt, T.M., (2008). Identifying methods to improve heart surgery: an operative approach and strategy for implementation on an organizational level. *Eur J Cardiothorac Surg*, 34, 1027–1033

Endsley, M.R., (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, 37, 32–64.

Flin, R., Yule, S., Paterson-Brown, S., Maran, N., Rowley, D., Youngson, G. G., (2007) Teaching surgeons about non-technical skills. *The Surgeon*, 5, 2, 86-89.

Hugh, T. B., (2002). New strategies to prevent laparoscopic bile duct injury - surgeons can learn from pilots. *Surgery*, 132, 826-835.

Karl, R.C., (2009). Aviation, *J Gastrointest Surg*, 13, 6–8.

Kenyon, T.A., Lenker ,MP, Bax, T.W., Swanstrom, L.L., (1997). Cost and benefit of the trained laparoscopic team. A comparative study of a designated nursing team vs a nontrained team. *Surg Endosc*, 11, 812–814.

Kitch, B.T., Cooper, J.B., Zapol, W.M., Marder, J.E., Karson, A., Hutter, M., Campbell, E.G., (2008), Handoffs causing patient harm: a survey of medical and surgical house staff. *Joint Commission Journal on Quality and Patient Safety*. 34, 563-570.

Lee, D.Y., Myers EA, Rehmani SS, et al, (2012) Surgical residents' perception of the 16-hour work day restriction: concern for negative impact on resident education and patient care. *J Am Coll Surg*. 215, 868-877

Makary, M., Holzmueller, C., Sexton, B., et al, (2006). Operating room debriefings. *Joint Commission Journal on Quality and Patient Safety*, 32, 407–410.

Makary, M., Holzmueller, C., Thompson, D., et al, (2006). Operating room briefings: working from the same page. *Joint Commission Journal on Quality and Patient Safety*, 32, 351–355.

Mitchell, L., Flin, R., Yule, S., Mitchell, J., Coutts, K., & Youngson, G., (2011). Thinking ahead of the surgeon. An interview study to identify scrub nurses' non-technical skills. *International Journal of Nursing Studies*, 48, 818–828.

Muller, S., Zalunardo, M.P., Hubner, M., Clavien, P.A., Demartines, N.A., (2009). Fixed team program reduces complications and length of hospital stay after open colonic surgery. *Gastroenterology*,136, 842–847.

Neily, J., Mills, P.D., Young-Xu, Y., et al, (2010). Association between implementation of a medical team training program and surgical mortality. *JAMA*, 304, 1693–1700.

Sexton, B., Thomas, E., & Helmreich, R., (2000). Error, stress and teamwork in medicine and aviation: cross sectional surveys. *British Medical Journal*. 320, 745–749.

Stepaniak, P.S., Heij, C., Buise, M.P., Mannaerts, G.H.H., Smulders, J. F., & Nienhuijs, S. W., (2012). Bariatric Surgery with Operating Room Teams that Stayed Fixed During the Day: A Multicenter Study Analyzing the Effects on Patient Outcomes, Teamwork and Safety Climate, and Procedure Duration. *Anesth Analg*, 115, 1384–1392.

Weaver, S.J., Rosen, M.A., Diaz-Granados, D., Lazzara, E.H., Lyons, R., Salas, E., et al, (2010), Does teamwork improve performance in the operating room? A multilevel evaluation. *Jt Comm J Qual Patient Saf*, 36,133–42

Wolf, F.A., Way, L.W., Stewart, L, (2010). The Efficacy of Medical Team Training: Improved Team Performance and Decreased Operating Room Delays. A Detailed Analysis of 4,863 Cases. *Ann Surg*, 252, 477-483.

Young-Xu, Y., Neily, J., Mills, P.D., et al, (2011). Association between implementation of a medical team training program and surgical morbidity. *Archives of surgery*, 146, 1368-1373.

Youngson, G.G., Flin, R., (2010) Patient safety in surgery: non-technical aspects of safe surgical performance. *Patient Saf Surg*, 18, 4, 4. doi: 10.1186/1754-9493-4-4.

Yule, S., Flin, R., Paterson-Brown, S. & Maran, N., (2006). Non-technical skills for surgeons in the operating room: A review of the literature. *Surgery*, 139, 140-149.

Agility in Allocating Decision-Making Rights For Cyberspace Operations

Steven W. STONE

Robert Morris University swsst320@mail.rmu.edu

ABSTRACT

In 2011, the United States (U.S.) Department of Defense (DoD) named cyberspace a new domain of warfare. The U.S. Cyber Command and the Military Services are working to make the cyberspace environment a suitable place for achieving national objectives and enabling military command and control (C2). The DoD's emerging cyberspace doctrine attempts to address the uniqueness of military operations in cyberspace and clarify the command relationships for cyberspace operations. However, military planners are attempting to apply C2 doctrine developed for military operations in the physical domain to military operations in the cyberspace domain. The spatial and temporal dimensions of cyberspace are significantly different than the physical domain and are significantly more complex and dynamic. This situation suggests a need to consider the relationship of the organization to its environment in order to determine the appropriate allocation of decision-making rights. This paper presents on-going research into the factors influencing agility in allocating decision-making rights for cyberspace operations.

KEYWORDS

Decision Making ; Command and Control ; Military ; Cyberspace Operations ;

INTRODUCTION

The growing use of cyberspace has reached the point where a wide range of social, political, informational, economic and military activities are dependent on it and are vulnerable to both interruption of its use and usurpation of its capabilities (Kuehl, 2009). The physical platforms, systems, and infrastructures that provide global connectivity to link information systems, networks, and human users with massive amounts of information that can be digitally sent anywhere, anytime to almost anyone has greatly increased access to information and has affected human cognition, dramatically impacting human behavior, and decision making (Kuehl, 2009).

In order to effectively conduct cyberspace operations in support of the Nation's security and military operations, the Secretary of Defense directed the establishment of U.S. Cyber Command in 2009 (United States Department of Defense, 2009). In 2011, the U.S. Department of Defense (DoD) named cyberspace a new domain of warfare (Williams, 2014). The purpose of both actions was to achieve the United States' national security objectives in or through cyberspace. In support of these objectives, the Under Secretary of Defense for Policy stated "There is a compelling need for a comprehensive, robust and articulate cyber power theory that describes, explains, and predicts how our nation should best use cyber power in support of U.S. national and security interests" (Kramer, Starr, & Wentz, 2009, p. xv). Subsequent to that statement, the U.S. military began to develop an understanding of and doctrine for utilizing cyberpower; "the ability to use cyberspace to create advantages and influence events in all of the operational environments and across the instruments of power" (Kuehl, 2009, p. 38).

The U.S. Cyber Command and the military services are working to make the cyberspace environment a suitable place for achieving our national objectives and enabling military command and control (C2). The DoD defines cyberspace as "A global domain within the information environment consisting of the interdependent network of information technology infrastructures and resident data, including the Internet, telecommunications networks, computer systems, and embedded processors and controllers" (United States Department of Defense, 2014, p. 63). The DoD further defines cyberspace operations as "The employment of cyberspace capabilities where the primary purpose is to achieve objectives in or through cyberspace" (United States Department of Defense, 2014, p. 63). In 2013, the DoD published

Joint Publication 3-12, Cyberspace Operations (U.S. Department of Defense, 2013). This emerging doctrine attempts to address the uniqueness of military operations in cyberspace and clarify cyberspace operations command relationships. However, there is a lack of research on decision-making in the face of the complex dynamics presented by the cyberspace domain.

STATEMENT OF THE PROBLEM

The problem facing the DoD is that it does not understand the factors affecting nor how to implement agility in allocating decision-making rights in the face of the complex dynamics presented by the cyberspace domain. The cyberspace domain is significantly different from the physical domain in both the temporal and spacial dimensions. Cyberspace is inherently global in nature and cyber effects often occur at the speed of light. This new domain presents a much more dynamic and complex operational environment for the U.S. military. Thus, military operations in cyberspace likely require different and more agile C2 and decision making methods to be successful. However, the DoD is currently applying C2 doctrine developed for operations in physical space to operations conducted in cyberspace.

This attempt by military planners to apply C2 doctrine developed for physical military operations to cyberspace operations is inappropriate. The temporal and spatial differences presented by cyberspace require the military to examine its long-held doctrine for C2. This situation suggests a need to consider the relationship of the organization to its environment in order to determine the appropriate organizational design (Galbraith, 1973, p. v). Several authors have called for additional research in this area. Alberts (2014) has called for research into the "...identification of key variables and relationships that should be included in a model of Command and Control Agility Potential whose output would be an entity's C2 AQ (agility quotient)" (Alberts, 2014, p. 1). Gore, Banks, Millward, & Kyriakidou (2006) conclude that a major goal of decision-making research is the development of ecologically valid practical methods for minimizing error and improving decision quality.

Differences in the Cyberspace Domain

For much of recorded history, military forces had only two physical domains in which to operate, the land and the sea. Both domains had different physical characteristics and humans used different technologies to operation in these domains. In addition to walking, military operations in the land domain were enhanced by the wheel, and various vehicles. Because humans can swim for only so long, war fighting on the sea was possible only with the aid of technology: the galley, sailing ship, steamship, and nuclear submarine (Kuehl, 2009). Two additional war-fighting domains were added in the 20th Century: air and outer space. Military operations in both of these domains were made possible by advances in technology, the development of aircraft and spacecraft. Each of these four physical domains is marked by radically different physical characteristics, and they are usable only through the use of technology to exploit those characteristics (Kuehl, 2009).

Cyberspace has uniquely defining characteristics when compared to the land, sea, air, and outer space domains. First, cyberspace is a man made domain. While the physical characteristics of cyberspace come from electromagnetic forces and phenomena that exist and occur in the natural world, cyberspace is a human designed environment, created to use and exploit information, human interaction, and intercommunication. Cyberspace was created not to sail the seas or orbit the earth, but rather to "create, store, modify, exchange, and exploit" information via electronic means (Kuehl, 2009). Human kind can capture any type of information, store that information as a string of bits and bytes, modify it to suit our purposes, and then transmit it instantly to every corner of the globe.

Second, cyberspace is global in nature. The effects of war fighting in the physical domains are typically limited to an easily identifiable geographic area. A bomb affects a small radius around its detonation point. A bullet affects a small area around its aim point. Cyberspace effects are not limited to a small local area. Cyber effects are often global in nature. For example, malware frequently infects computer systems worldwide.

Third, activities in cyberspace can happen extremely rapidly. As cyberspace is created using electromagnetic forces found in nature, effects in cyberspace can travel at the speed of light. Kuehl states, "What makes cyberspace neither aerospace nor outer space is the use of the electromagnetic spectrum as the means of "movement" within the domain, and this clear distinction from other physical environments may be crucial to its further development within the national security structure" (Kuehl, 2009, p. 31).

Fourth, cyberspace is incredibly complicated, comprising millions of separate hardware devices, running software with millions of potential settings, and processing millions of bits of data. Modern operating systems have thousands of settings. Many network security devices have hundreds of thousands of rules running on them at any point in time. Richard Hale, the DoD's Chief Information Security Officer states, "No human being can understand this. There is no way any human analyst has a prayer of taking all of thousands of settings multiplied by thousands of settings and making sense of that" (Hale, 2014).

Fifth, unlike the physical domains, where nature often sets the conditions of the environment, many decisions regarding the behavior of cyberspace are made by the software running on those devices. Conducting operations in cyberspace is done by changing the configuration of these complicated pieces of equipment. Peter Fonash, Chief Technology Officer for the Department of homeland Security Office of Cybersecurity and Communications, states, "The first technology that I would want to have is a capability to do automated decision-making and automated courses of action. Instead of waiting for a human to perceive a threat, make sense of it, and decide on a response — let alone wait for higher-ups to authorize it — we need software that can perform all those functions by itself, moving at the same speed as the attacking malware" (Fonash, 2014)

DECISION MAKING THEORY

Hoffman describes organizational design as "the relatively enduring allocation of work roles and administrative mechanisms that creates a pattern of interrelated work activities and allows the organizations to conduct, coordinate, and control its work activities" (Hoffman, 1998, p.6). One of the primary dimensions of organizational design is the decision making structure. Hoffman states that the "Decision making structure involves the centralization and decentralization of decision making. Organizational decision-making has been formally defined as being the process of identifying and solving problems within organizations (Hoffman, 1988, p. 7). The performance of an organization is determined, at least partially, by how well problems are identified and solved. Thus, an organization's decision-making structure is one of the most critical areas of the organization's design. Several theoretical models of decision-making, such as Albert's and Hayes' model of C2, the Military Decision Making Process, Galbraith's Information Processing Model, Klein's Recognition-Primed Decision Model and Shattuck & Miller's Dynamic Model of Situated Cognition suggest factors that may affect the allocation of decision-making rights for cyberspace operations.

Theoretical Model of Command and Control and Decision Making Agility

The U.S. Military's C2 doctrine has been developed and refined over many years of military operations in the industrial age. However, there is significant debate as to whether these decision-making relationships will be effective in the information age. Alberts argues that the traditional DoD C2 approach is no longer sufficient for military operations in cyberspace.

Alberts and Hayes (2006) describe three dimensions of a theoretical model of C2 or, in civilian parlance, organizational culture, that are useful in this research: The organization's allocation of decision-making rights, the organization's patterns of interaction, and the organization's distribution of information. The allocation of decision rights is a linear dimension with two logical endpoints. At the origin of the allocation of decision-making rights on the horizontal axis, decision-making rights are unitary, all the rights held by a single actor. At the other end axis, decision-making rights are allocated uniformly with every entity having equal rights in every decision (Alberts & Hayes, 2006). Alberts and Hayes hypothesize that complex dynamic environments, like cyberspace operations, require more agile approaches to C2. Alberts' hypothesis is that agile C2 requires the organizational ability to rapidly change their approach towards each of the three variables in the theoretical model of C2 (Alberts & Hayes, 2006).

Military Command & Control and The Military Decision Making Process

The U.S. DoD has a large body of organizational design documentation that describes how the U.S. military is organized and functions. In military parlance this body of documentation is called doctrine. The U.S. military's term to describe its organizational design and decision-making process is command and control (C2). The DoD defines C2 as "The exercise of authority and direction by a properly designated commander over assigned and attached forces in the accomplishment of the mission" (United States Department of Defense, 2014, p. 44).

Information Processing Theory

Galbraith's Information Processing Theory presents a framework to describe the relationship of an organization to the information environment it faces (Galbraith, 1973; Galbriath, 1974). Galbraith states that the basis of his Information Processing Theory is "... the greater the task uncertainty, the greater the amount of information that must be processed among decision makers during task execution in order to achieve a given level of performance" (Galbriath, 1973, p. 4). Galbraith also states that the type of information processed, either quantitative or qualitative, affects where the information should be processed. This theory is very applicable to the military's cyberspace operations C2 issue.

Naturalistic Decision Making

Naturalistic Decision Making (NDM) provides a theory and methodology to describe how decision makers actually make decisions in complex domains. NDM research focuses on what decision makers actually do in fast-paced, complex, and dangerous situations where there is not time to perform elaborate evaluation of alternatives or to optimize the decision (Lipshitz, Klein, & Carroll, 2006). This theoretical framework describes the environmental characteristics applicable to NDM as: ill-structured problems; uncertain, dynamic environments; shifting, ill-defined or competing goals; action/feedback loops; time stress and high stakes; organizational goals and norms (Orasanu and Connolly 1993). Two, NDM based, decision-making theories provide additional insight in to potential factors affecting cyberspace operations decision-making: The Recognition Primed Decision (RPD) model and the Dynamic Model of Situated Cognition (DMSC).

PURPOSE OF THE STUDY

The purpose of this quantitative exploratory study is to identify the factors influencing the U.S. Military's agility in allocating decision-making rights for cyberspace operations. This study will analyze factors identified from the literature and factors identified by experts in the field. The goal of this study is to provide military decision makers with a list of factors to consider when determining the allocation of decision-making rights for cyberspace operations.

RESEARCH QUESTION

The research question for this study is: What factors influence the U.S. Military's agility in allocating decision-making the rights for cyberspace operations?

METHODOLOGICAL DESIGN

Given the complex nature of this problem and the somewhat open ended nature of the research question, the researcher proposes to use the Delphi research method to identify the factors influencing the U.S.

Military's agility in allocating decision-making rights for cyberspace operations.

The Delphi panel will be recruited from experts in C2 and Cyberspace Operations. For purposes of this research, an expert is defined as a person that has at least five years of practical experience working in cyber operations; or a person that has an advanced degree in an information management field with over 10 years of research in cyberspace operations, C2, decision-making theory, teaching, publication experience; or a combination of the two. The panel will be recruited from the C2 research community through the researcher's participation in and contacts with International Command and Control Research and Technology Symposium (ICCRTS), the Naturalistic Decision Manking (NDM) Conference, and the Military Cyberspace Professional Association (MCPA).

SIGNIFICANCE OF THE STUDY

This research will add to the body of knowledge in that it will assist the U.S. military to define the C2 structures and procedures that will enable them to be successful in conducting cyberspace operations. The military officers and civilians leading cyberspace operations have been influenced by their military education and experience in leading military operations in the physical space. As such, they are attempting to describe how they will conduct cyberspace operations using the concepts and doctrine from physical operations. This approach is flawed because the time and space characteristics of cyberspace are significantly different than the physical domain. Therefore, research into new approaches to C2 is necessary to achieve success in cyberspace operations.

SUMMARY

As discussed in this paper, the U.S. military is facing challenges in cyberspace that present a much different environment than operations in the physical space. The temporal and spatial differences presented by cyberspace require the military to examine it long-held doctrine for C2. Albert's and Hayes' model of C2, the Military Decision Making Process, Galbraith's Information Processing Model, Klein's Recognition-Primed Decision Model and Shattuck & Miller's Dynamic Model of Situated Cognition provide the theoretical framework to examine the factors influencing the allocating decision-making the rights for cyberspace operations. The outcome of this study will provide military decision makers with a list of factors to consider when determining the allocation of decision-making rights.

REFERENCES

Alberts, D. S., & Hayes, R. E. (2006). *Understanding Command and Control.* Washington DC: Office Of The Assistant Secretary Of Defense For Networks And Information Integration, Command Control Research Program. Retrieved from http://www.dodccrp.org/files/Alberts_UC2.pdf.

Alberts, D. S. (2007). *Agility, Focus, and Convergence: The Future of Command and Control.* Washington DC: Office Of The Assistant Secretary Of Defense For Networks And Information Integration, Command Control Research Program. Retrieved from http://www.dodccrp.org/html4/journal_main.html.

Fonash, P. (2014, December). Public comments in R. Rodriguez (Chair). *SINET Showcase and Workshops 2014.* Symposium conducted in Washington D.C.

Freedberg, S.J., (2014). Moving mountains in cyber war: Automated virtual 'maneuver'. *Breaking Defense.* Retrieved from: http://breakingdefense.com/2014/12/moving-mountains-in-cyber-war-automated-virtualmaneuver.

Galbraith, J. R. (1973). Designing complex organizations. Reading, Massachusetts: Addison-Wesley Publishing Co., Inc.

Galbraith, J. (1974). Organization design: an information processing view. *Interfaces, 4(3)*, 28-36.

Gore, J., Banks, A., Millward, L., & Kyriakidou, O. (2006). Naturalistic decision making and organizations: Reviewing pragmatic science. *Organization Studies*, 27(7), 925-942.

Grisham, T. (2009). The delphi technique: A method for testing complex and multifaceted topics.

International Journal of Managing Projects in Business, 2(1), 112-130. doi:http://dx.doi.org/10.1108/17538370910930545

Hale, R. (2014, December). Public comments in R. Rodriguez (Chair). *SINET Showcase and Workshops 2014*. Symposium conducted in Washington D.C.

Hoffman, J. (1988). *The effects of strategic and operational decision making structure on organizational performance: Technology as a moderator* (Doctoral Dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8818630).

Keeney, S., Hasson, F., & McKenna, H. (2011). The Delphi Technique in Nursing and Health Research. Oxford, United Kingdom: Wiley-Blackwell.

Kramer, F. D., Wentz, L.K. & Starr, S. H. (Eds.). (2009). *Cyberpower and national security*. Dulles, VA: Potomac Books, Inc.

Kuehl, D.T., (2009), From cyberspace to cyberpower: Defining the problem. In Kramer, F. D., Wentz, L.K. & Starr, S. H. (Eds.), *Cyberpower and national security* (pp. 24-42). Dulles, VA: Potomac Books, Inc.

Linstone, H. A., & Turoff, M. (Eds.). (1975). *The Delphi method: Techniques and applications* (Vol. 29). Reading, MA: Addison-Wesley.

U.S. Department of Defense. (June 23, 2009). *Establishment of a subordinate unified U.S. cyber command under*

U.S. strategic command for military cyberspace operations. Retrieved from: http://online.wsj.com/public/resources/documents/OSD05914.pdf.

U.S. Department of Defense. (2011). *Department of Defense (DoD) Information Technology (IT) Enterprise Strategy* www.dtic.mil/doctrine/new_pubs/jp3_12R.pdf.

U.S. Department of Defense. (2014). *Joint Publication 1-02: Department of Defense Dictionary of Military and Associated Terms*. Retrieved from: http://www.dtic.mil/doctrine/new_pubs/jp1_02.pdf.

Williams, B.T. (2014). The joint force commander's guide to cyberspace operations. *Joint Force Quarterly, 73,* 12.

A Naturalistic Decision Making Perspective on Food Purchasing and Resource Allocation Decisions Among Low-income Households

Drew A. Zachary[a]
[a]*Brandeis University, Heller School for Social Policy and Management*

ABSTRACT
Low-income households' food purchasing decisions provide a real-world example of resource allocation under extreme constraints. This paper examines such food purchasing as a naturalistic decision making (NDM) process from the perspective of Naturalistic Decision Theory, by analyzing food purchasing as a process in which individuals use expert knowledge of their food environment and of their extreme economic constraints to develop and apply rational strategies, and balance competing goals related to food. The health and nutrition problems faced by low-income families have motivated creation of public programs to provide food-purchasing assistance, but the design characteristics of the programs create new constraints that act as additional constraints on food purchasing decisions. This paper extends NDT to a new non-technical domain, contributes to theoretical understanding of how low-income people make routine decisions under extreme constraints, and points out how NDT can be used to inform policy development and refinement.

KEYWORDS
Practical application; theory and modeling; policy; cognitive field research and CTA; decision making; expertise

INTRODUCTION

This paper describes the underlying context-dependent processes associated with food purchasing decisions. Food purchasing is analyzed as a naturalistic decision making (NDM) process from the perspective of Naturalistic Decision Theory (NDT) [Klein & Klinger, 1991; Klein, 1993]. The paper applies NDT to frame the ways in which severely resource constrained individuals make decisions about food purchasing.

Low-income households' food purchasing decisions provide a real-world example of resource allocation under extreme constraints. Food purchasing is a recurring decision that allows people to develop expertise as they gain experience and knowledge about the task and task environment. The task occurs in the context of a local environment, and presents a complex problem: purchase food with competing needs, goals, and constraints (for example, balancing food quantity and dietary quality). People living in low-income communities face constraints such as distance from supermarkets and quality of produce and other fresh foods [Morland, Diez Roux & Wing, 2006; Woolf & Braveman, 2011], limited access to affordable healthy foods, and a relatively high presence of unhealthy, low-cost foods [Cassady, Jetter & Culp, 2007; Bartlett et al., 2013; LaVeist et al, 2011; Moore & Diez-Roux, 2006; Palmer et al., 2007; Raja, Ma & Yadav, 2008; Treuhaft & Karpyn, 2010]. Thus, food purchasing among low-income households provides an opportunity to study how people use detailed knowledge of their environment to make resource allocation decisions, and how qualities of this environment influence decision making about frequent behaviors. In particular, low-income households receiving Supplemental Nutrition Assistance Program (SNAP, formerly Food Stamps) present an ideal example, because SNAP poses specific constraints. Many low-income families receive SNAP benefits as a main resource for food purchasing. Most states issue benefits over a period of several days[11], creating a monthly "SNAP cycle"[12] in low-income communities with a high concentration of SNAP recipients. Prior studies indicate that this temporal constraint influences food purchasing decision processes by leading to food purchasing once per month, corresponding to SNAP benefit transfers

---

[11] For example, in Maryland, benefits are disbursed from the 6th to the 15th of each month in alphabetic order, in Maryland, with last names A-B on the 6th and so on.

[12] The "SNAP cycle" refers to the one-month period between SNAP electronic benefit transfers, although it does not typically coincide with the start of the calendar month.

[Kharmats et al., 2014; US Department of Agriculture, 2014; Zachary, Palmer, Beckham & Surkan, 2013]. Prior research also indicates that low-income individuals strategize about how to pay for foods at different times of month, at different food stores, and with various food purchasing resources [Clifton, 2004; Rose, 2011; USDA, 2014; Zachary Palmer & Surkan 2012; Zachary, Palmer, Beckham, & Surkan, 2013].

Naturalistic Decision Theory (NDT) has been applied primarily to the decisions/behavior of technical experts, explaining the implicit processes used by experts in making real-time decisions. However, everyday people (e.g. grocery shoppers) also make routine decisions that allow them to gain expertise and apply underlying decision process to improve the efficiency of the task from their perspective. This is an area where NDT could be applied to enhance the theoretical understanding of everyday economic decisions/behaviors. Moreover, low-income people with greater constraints likely develop (out of necessity) more sophisticated underlying processes for allocating resources.

APPLICATION

Problem environment and context

Food purchasing requires low-income individuals to balance competing goals:

- Consistently provide enough food for members of the household;
- Provide healthy food if possible, particularly foods that meet needs of household members with health conditions; and
- Accommodate preferences of household members (especially children) [Pollard et al., 2014; Wingert et al., 2014; Zachary et al., 2013].

However, individuals must do so in an extremely constrained environment. *Resource constraints* include very limited income: for example, approximately 85% of households receiving SNAP have income below the federal poverty level [Creswell, 2007]; and the average monthly benefit per person in FY2012 was only $133 [Glanz & Yaroch, 2004]. There are also restrictions on how some food purchasing resources can be used. Many low-income households receive or qualify for programs in addition to SNAP such as WIC[13] and free/reduced price school breakfast and lunch. These resources are more restricted in that they can only be redeemed for specific foods or are delivered in-kind, whereas SNAP can be redeemed for almost any food product at approved grocery or food stores. Additionally, the household might not have access to a car, making transportation expensive and/or time consuming.

Low-income households also face *time constraints* on their purchasing decisions. Individuals have a limited amount of time to complete the task, and they must consider time cost of purchasing trips and traveling to various food stores, transportation cost, and the time/distance they will have to travel to food stores. They also face the tradeoffs of accessibility and desirability, in that unhealthier foods might be available within a shorter distance and in greater supply, but might have comparatively higher prices. Or, that a food store has a wider selection of products or better pricing, but with great travel cost [Zachary, Palmer & Surkan 2012; Zachary, Palmer, Beckham & Surkan, 2013]. Food purchasers may also be distracted from the task if they must bring children along to food stores [Wingert et al., 2014]. An additional constraint on SNAP recipients in particular is that SNAP is only disbursed to households once per month.

Food purchasing also involves *uncertainty*: unreliable quality introduces financial risk that non-perishable foods will spoil and thus waste resources [Zachary, Palmer & Surkan 2012; Zachary, Palmer, Beckham & Surkan, 2013].

DECISION THEORY CONTEXT

Naturalistic Decision Theory (NDT)

NDT [Klein & Klinger, 1991; Klein, 1993; Riegel Dickson & Topaz, 2013] emerged as an empirically based alternative to formally rational and experiment-based behavioral theories of decision-making. Rational choice theory models decision making as a process of comparing options to reach an optimal outcome [Klein, 1993; Sunstein &

---

[13] The Supplemental Nutrition Program for Women, Infants, and Children (WIC) provides vouchers for foods such as whole grains, dairy products, fruits and vegetables. WIC also includes nutritional counseling and in-store shelf labeling to identify approved healthy products [Zachary, Palmer & Surkan 2012].

Thaler, 2008; Zachary & Ryder, 1997]. This is normative in that it assumes there is an optimal decision based on a predicted ideal outcome. Yet this process of optimization does not reflect how humans actually make decisions [Klein, 1993; Klein & Klinger, 1991; Sunstein & Thaler, 2008; Zachary & Ryder, 1997].

NDT explains problem solving or decision-making processes as fundamentally situated in a real world (i.e. "naturalistic") context. NDT is based on the empirical concept that actual human behavior does not "fit into a decision tree framework" [Klein & Klinger, 1991; Zachary & Ryder, 1997]. Rather than comparing decisions to theoretical optimal outcomes, NDT explains that decisions make sense in the context of a naturally encountered situation. The task situation/context are important to the actor in "framing the problem," and thus people make decisions differently in real world settings than in laboratory settings [Klein, 1993; Zachary & Ryder, 1997]. From this theoretical perspective, health behaviors can be analyzed as decision processes that occur and make sense in the context of an environment [Riegel Dickson & Topaz, 2013]. Whereas prior decision theories explained decision-making as a process used to evaluate a hierarchy of alternatives [Luce & Raiffa, 1989], NDT instead explains how "people make effective decisions without performing analyses" [Klein, 1993; Klein & Klinger, 1991]. Thus, rather than explaining unhealthy purchases or purchasing strategies as irrational or biased, NDT allows us to explain them as outcomes of a rational process that occurs in the context of the task environment, and as a result of how people think about and apply their expertise to the task of purchasing food.

NDT explains that some things we consider decisions are only formal decisions in retrospect, and in real time they are made through a more implicit process [Kahneman, 2011]. As people gain expertise in a task and task environment, the process becomes more advanced and automatic, and individuals can apply strategies that are better tailored to the situation. Experts recognize situations and know what decision to make based on prior experience and knowledge [VanLehn, 1996].

This theoretical framework links micro (individual level) cognition and actions with macro-level influences and patterns. The concept of cognitive skill acquisition explains that people learn to complete tasks by gaining experience [VanLehn, 1996]. Although some decisions are made at the individual level, decision-making is context dependent, and individuals who share environments, relevant goals, and similar experience will tend to approach problems similarly [VanLehn, 1996]. Although each person's decision process will not be exactly the same, people complete tasks using a common set of strategies with predictable variability [Klein, 1993; VanLehn, 1996; Zachary & Ryder, 1997]. By studying the individual processes used by people under similar conditions, it is possible to understand how people think about the task and decision environment, and to identify the common strategies used [VanLehn, 1996].

To date, NDT and cognitive skill acquisition theory have been applied largely to tasks gained through professional experience, in which people apply technical expertise. They have not yet been applied in in everyday tasks that present complex resource allocation problems. Applying a NDT perspective, food purchasing can be viewed as a task or process in which people apply expertise. For example, low-income shoppers do not make a formal calculation of the cost of time, gas, etc. for all possible alternatives when considering at which store to shop. Rather, they will likely know where to go without using an explicit decision process, based on their experience buying food in the local community and their knowledge about relevant qualities of food stores. A shopper might choose a supermarket that has the lowest prices or is the most efficient in terms of transportation cost; or by considering factors such as distance, overall price levels, perceived quality and selection of food at the store relative to other accessible stores, all without a conscious decision process. Cognitive skill acquisition points out that, because of its frequent re-occurrence, the process of purchasing food provides an opportunity for the decision maker to learn recurring aspects of purchasing strategies that meet their need, those that do not, and their major points of difference. As individuals gain experience, the process evolves as one involving more expert, but more implicit, decisions. Study these implicit processes can provide insight into how low-income individuals with similar environmental constraints approach the task of allocating food purchasing resources

Behavioral decision theories

Some aspects of behavioral decision theories provide useful ideas for analyzing decision making within an NDT framework – for example, systematic biases resulting in suboptimal outcomes, satisficing, and heuristics [Ariely, 2008; Kahneman, 2001; March, 1978; Sunstein & Thaler, 2008; Tversky & Kahneman, 1974]. In contrast with rational choice decision models, which reflect an unbounded field of alternatives, behavioral decision theories adopt the perspective that in reality decision-making reflects a bounded or constrained view of possible alternatives – a concept referred to as "bounded rationality." Within this constrained set of alternatives, people "satisfice" or choose a "good enough" option even if it is suboptimal [March, 1978; Tversky & Kahneman, 1974]. Moreover, due to cognitive biases, people make decisions in predictably sub-optimal ways (compared to the optimal choice that normative decision theory would prescribe) [Ariely, 2008; Simon, 1996; Sunstein & Thaler, 2008; Tversky &

Kahneman, 1974]. Some biases are related to the external environment. People do not simply determine the rational decision and then execute it; rather, choices are systematically biased by the way environments are designed (e.g. choosing default options against one's best interest, loss aversion, etc.) [Ariely, 2008; March 1978; Sunstein & Thaler, 2008; Tversky & Kahneman, 1974].

Behavioral decision theories have two critical limitations for studying real-world behavior. First, they were developed based on research conducted in laboratory settings, using problems that people would not face in the real world, rather than studying actual decision processes and outcomes in context [Klein & Klinger, 1991; Zachary & Ryder, 1997]. Second, behavioral theories are often not based on studies involving people with expertise in the problem and environment. As a result, the theories "over-formalize" decision-making, as compared to the less formal processes used in real situations, especially by experts [Klein 1993; VanLehn, 1996; Zachary & Ryder, 1997].

These behavioral economic (BE) concepts have been criticized for creating a false dichotomy between a "right" and "wrong" decision, where the wrong decision is determined based on a theoretical optimal outcome [Klein & Klinger, 1991; Zachary & Ryder, 1997]. However, they are useful for analyzing how people make decisions in which there is an ideal choice. For example, with respect to dietary quality or health conditions, there are in fact more and less healthy food choices. Behavioral decision theories explain that people will predictably make unhealthy choices, based on environmental influences and other competing goals and constraints [Ariely, 2008:]. Thus, they are useful for studying the ways in which available resources and aspects of environmental context (e.g. the monthly SNAP cycle) influence purchasing and resource allocation decisions, and the ways these constraints affect consideration of dietary quality. BE concepts provide insights about the cognitive tools used to make food purchasing decisions within time constraints, or to weigh cost-effectiveness against healthfulness as a criterion.

NDT ACCOUNT OF FOOD PURCHASING AMONG LOW-INCOME HOUSEHOLDS

Food purchasing decision processes among low-income shoppers

There is an emerging body of research on decision processes and strategies for food purchasing used by low-income shoppers. To elaborate the relationship between individual and environmental qualities, research has examined low-income shopper's food purchasing decisions in the context of their environments. Recent studies have identified common shopping strategies used by low-income consumers with limited healthy food access, including consolidating shopping trips to limit transportation cost, purchasing non-perishable items in bulk, avoiding perishable foods that are likely to spoil, and identifying affordable products using sale ads and in-store sale labeling [Clifton 2004; Rose, 2011; Zachary, Palmer, Beckham & Surkan, 2013; Zenk et al. 2011]. A Baltimore-based qualitative study of food purchasing decision processes found that shoppers' primary goal was to provide enough food for the household. Shoppers developed decision criteria to identify affordable purchases, which in the context of their community environment and immediate shopping environment, led them to buy more unhealthy and fewer healthy groceries than they would prefer [Zachary, Palmer, Beckham & Surkan, 2013]. Prior research findings suggest that low-income shoppers with children obtain food with monetary resources such as cash, SNAP, and WIC, and non-monetary resources such as school meals and food pantries.

A recent study examined food purchasing strategies used by SNAP recipients [USDA, 2014]. 90 SNAP households with children were interviewed about food security and how they cope with changes in resource levels that make it difficult to afford food. The authors identified strategies that households use to provide enough food in the face of financial hardship, including parents restricting their intake to save food for children, turning to social networks for assistance if possible, and carefully planning shopping trips and purchases. Although this study listed general "coping strategies" for dealing with poverty and very limited food budgets, it did not identify decision processes used in purchasing food with SNAP benefits.

Several studies demonstrate that monthly benefit transfers led SNAP recipients to make one main grocery-purchasing trip per month. One study found that, in low-income communities with large concentrations of SNAP recipients, the days of the month on which SNAP benefits are transferred coincide with crowding and higher prices at food stores [Zachary, Palmer, & Surkan, 2012], which places an additional environment constraints on decision making. Several studies have described participants running out of money at the end of the SNAP cycle, making it difficult to provide food for their households [Kharmats et al. 2014; Leung et al, 2013; Zachary, Palmer, & Surkan, 2012; Zachary, Palmer, Beckham & Surkan, 2013]. One of these recent studies examined dietary quality among SNAP recipients at different points during the one-month period after SNAP transfers and found that time since SNAP transfer had a significant effect on dietary quality [Kharmats et al, 2014]. However few studies have specifically analyzed the influence of this cycle on individuals' decision-making. Additional research is needed to further examine the relationship between SNAP benefit transfers and purchasing decisions, especially in the context of other environmental influences on purchasing.

Decision influences and context

Prior studies have identified specific qualities of local food environments that influence food purchasing. Low-income communities form the environmental context for many SNAP recipients' food purchasing decisions. Many low-income urban areas have few if any high quality grocery stores, full service supermarkets, or availability of fresh food, but a relatively high presence of unhealthy food options such as corner stores and fast food restaurants [Cassady; Jetter & Culp 2007; Dutko, VerPloeg & Farrigan 2012; LaVeist et al. 2011; Moore & Diez-Roux, 2006; Palmer et al 2007; Raja, Ma & Yadav, 2008; Treuhaft & Karpyn 2010]. Research demonstrates that such food environments encourage unhealthy food choices, limit residents' ability to eat healthfully [Franco et al, 2008; LaVeist et al. 2011; Moore & Diez-Roux, 2006; Moore et al., 2008; Powell et al., 2007; Treuhaft & Karpyn 2010]. Prior studies have described how structural qualities of the external environment act as barriers to healthy eating [Chang et al. 2008; Eikenberry & Smith, 2004; Fulp, McManus & Johnson, 2009; Glanz & Yaroch 2004; Monsivais, Agarwal, & Drewnowski; Story et al., 2008; Zenk et al., 2011]. Barriers include limited access to healthy foods, the high cost of healthy foods, and limited information about healthy eating.

Several experimental studies have demonstrated that manipulations to funding sources (financial incentives) can influence purchasing outcomes, but did not provide any data on the decision processes that led up to those purchases [Bartlett et al. 2013; Briggs et al, 2010; Fair Food Network, 2012; Hardin Fanning & Gokun, 2014; USDA, 2013]._Foster and colleagues conducted a cluster randomized control trial of an in-store marketing intervention to promote healthy purchasing, targeting low-fat dairy and healthier cereal, frozen meals, and beverages [Foster et al., 2014]. The authors randomized 8 supermarkets to receive changes to the store environment or no intervention and found that after 6 months, the treatment stores had greater sales of certain healthy products compared to the control stores [Foster et al., 2014]. Other studies indicate that qualities of food store environments influence food purchasing decisions, through shelf-labeling, perceived/visible quality of food, sale signs, and store layout [Foster et al., 2014; Gittelsohn et al., 2006; Wingert et al 2014; Zachary, Palmer, Beckham & Surkan, 2014]. More data are needed to understand how and why low-income shoppers make these decisions.

Many of these studies of low-income urban food environments are at an aggregate rather than an individual level. Those that do focus on the individual-level often use an ecological model that does not describe the relationships among influences or specifically how they influence decision-making.

Specific applications of NDT to food purchasing decisions among low-income households include the following:

- Food purchasing is a context-dependent process. Individuals in similar environments with similar constraints and experience in the task environment will approach the task in similar way(s). Based on NDT, low-income individuals will use various food purchasing resources in systematic ways. Resources will have different significance and qualities to the decision maker.
- Low-income individuals apply expertise and use an underlying rational decision process in making SNAP spending (and resource allocation) decisions. There will be a widely shared process, across the sample or across households with similar needs/constraints.
- Low-income individuals use heuristics to aid in decision making, satisfice within the options available to them, and make choices based on their needs and goals that are feasible within their given constraints.

CONCLUSION

This paper extends NDT to a new non-technical domain and explains its usefulness in that domain. This application contributes to the theoretical understanding of how low-income people (as experts in their local environments) make routine decisions under extreme constraints. This application focuses in particular on situations in which there is not a discrete single decision to be made, but rather a complex resource allocation problem. Using NDT to analyze such situations challenges the more common frame of behavioral and normative decision theories for analyzing purchasing. Whereas extant behavioral economic and normative decision theories would view certain food purchasing decisions or processes as irrational, a Naturalistic Decision Making lens would allow the possibility that they are rational and make sense in the context of the local environment, from the perspective of the decision maker. Studies applying NDT to analyze and model purchasing behavior can build on behavioral decision theory by providing *in situ* data on how consumer decision-making occurs in real world context, which complements *ex situ* behavioral economic research.

Future research
There has been limited empirical research on how individuals make decisions about spending their benefits. Prior studies provide some theoretical understanding of how low-income individuals develop strategies to purchase food for their households in the context of their local community and food store environments. Some studies have identified general decision strategies such as planning trips to minimize transportation cost. Other studies analyzed aggregate data on purchasing patterns and links to community characteristics, but relied on untested assumptions about how those links reflect individual decision-making [Bartlett et al., 2013; Fair Food Network, 2012; USDA, 2013]. However, few prior studies have examined how the SNAP cycle affects decision making over the course of a month, and how people solve the problem of acquiring food for the household under these constraints. Building on prior analyses that described a general process, future studies should aim to provide insights about decision making through a more detailed, intra-process analysis in the context of particular resources and constraints.

Specifically, further research should examine how temporal constraints on resource availability shape the food purchasing decision process, and how the availability of various resources affects the food purchasing/acquisition process. These questions could be operationalized in the context of food purchasing decisions of SNAP recipients.

Policy relevance of NDM/NDT
NDT is an important framework for developing evidence based policy design, because it sheds light on how people make complex decisions in real-world settings. It is important for policy designers to understand how individuals complete tasks and the ways in which policies, as part of environmental context, influence decision-making – especially with respect to behaviors policy aims to promote, such as purchasing specific healthy foods. Thus, studying naturalistic decision processes used in resource allocation/ food purchasing has the potential to identify opportunities for policy to promote dietary quality through SNAP-purchasing decisions. Other domains can also adopt this framework for developing theoretical understanding of routine behaviors in order to inform policy design that takes NDM processes into consideration, and thus serves the public more effectively.

ACKNOWLEDGMENTS
The author gratefully acknowledges Pamela Surkan, Jon Chilingerian, Lindsay Rosenfeld, Dolores Acevedo-Garcia, and Wayne Zachary for their insightful feedback on this work.

REFERENCES
Ariely, D. (2008). *Predictably irrational*. New York: Harper Collins.
Bartlett S., Klerman J., Wilde P., Olsho L., Logan C., Enver A. (2013). Healthy Incentives Pilot (HIP) Interim Report. U.S. Department of Agriculture, Food and Nutrition Service, Office of Research and Analysis, Project Officer: Danielle Berman, Alexandria, VA: July 2013. Retrieved from: http://www.abtassociates.com/Reports/2013/Healthy-Incentives-Pilot-(HIP)-Interim-Report.aspx
Briggs, S., Fisher, A., Lott, M., Miller, S., & Tessman, N., (2010). Real Food, Real Choice: Connecting SNAP Recipients with Farmers Markets, Community Food Security Coalition, Farmers Market coalition, June 2010.
Cassady, D., Jetter, K. M., & Culp, J. (2007). Is price a barrier to eating more fruits and vegetables for low-income families? *Journal of the American Dietetic Association, 107*(11), 1909-1915. doi: 10.1016/j.jada.2007.08.015
Chang, M. W., Nitzke, S., Guilford, E., Adair, C. H., & Hazard, D. L. (2008). Motivators and barriers to healthful eating and physical activity among low-income overweight and obese mothers. *Journal of the American Dietetic Assoc, 108*(6), 1023-1028. doi:10.1016/j.jada.2008.03.004
Clifton, K.J. (2004). Mobility strategies and food shopping for low-income families: A case study. *Journal of Planning Education and Research, 23*(4): 402-413. doi: 10.1177/0739456X04264919
Creswell, J.W. 2007. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*, Second ed. Thousand Oaks, CA: Sage.
Dutko, P., Ver Ploeg, M., Farrigan, T. (2012). Characteristics and influences factors of food deserts. Washington DC. US Department of Agriculture, Economic Research Service, Economic Research Report, No. 140.
Eikenberry, N., & Smith, C. (2004). Healthful eating: Perceptions, motivations, barriers, and promoters in low-income Minnesota communities. *Journal of the American Dietetic Association, 104*(7), 1158-1161. doi: 10.1016/j.jada.2004.04.023
Fair Food Network, 2012. Double Up Food Bucks: how it works. http://www.doubleupfoodbucks.org/how-it-works
Foster GD, Karpyn A, Wojtanowski, AC, Davis E, Weiss S, Brensinger C, Tierney A, Guo W, Brown J, Spross C, Leuchten D, Burns P, & Glanz K (2014). Placement and promotion strategies to increase sales of healthier

products in supermarkets in low-income, ethnically diverse neighborhoods: a randomized controlled trial. *American Journal of Clinical Nutrition, 99*(6):1359-1368. doi: 10.3945/ajcn.113.075572

Franco, M., Diez Roux, A.V., Glass, T.A., Caballero, B., & Brancati, F.L. (2008). Neighborhood characteristics and availability of healthy foods in Baltimore. American *Journal of Preventive Medicine, 35*: 561-567.

Fulp, R. S., McManus, K. D., & Johnson, P. A. (2009). Barriers to purchasing foods for a high-quality, healthy diet in a low-income African American community. *Family & Community Health. 32*(3), 206-217. doi: 10.1097/FCH.0b013e3181ab3b1d

Glanz, K., & Yaroch, A. L. (2004). Strategies for increasing fruit and vegetable intake in grocery stores and communities: Policy, pricing, and environmental change. *Preventive Medicine*, *39,* S75-S80. doi:10.1016/j.ypmed.2004.01.004

Kahneman, D. (2011). *Thinking Fast and Slow.* New York: Farrar, Straus and Giroux.

Kharmats A, Jones-Smith JC, Cheah YS, Budd N, Cuccia A, Mui Y, Trude A, Gittelsohn J. (2014). Relation between the Supplemental Nutrition Assistance Program cycle and dietary quality in low-income African Americans in Baltimore, Maryland. *American Journal of Clinical Nutrition, 99*(5): 1006-14. doi: 10.3945/ajcn.113.075994

Klein, G., Klinger, D. (1991). Naturalistic decision making. *Human Systems IAC Gateway, 11*(3): 16-19.

Klein, G.A. (1993). A recognition-primed decision (RPD) model of rapid decision making. *Decision making in action: Models and methods*, *5*(4), 138-147.

LaVeist, T., Pollack, K., Thorpe Jr, R., Fesahazion, R., and Gaskin, D. (2011). Place, not race: Disparities dissipate in southwest Baltimore when blacks and whites live under similar conditions. *Health Affairs, 30*(10), 1880-1887.

Leung C, Hoffnagle EE, Lindsay AC, Lofink HE, Hoffman VA, Turrell S, Willett WC, Blumenthal SJ. 2013. A qualitative study of diverse experts' views about barriers and strategies to improve the diets and health of Supplemental Nutrition Assistance Program (SNAP) beneficiaries. *Journal of the Academy of Nutrition and Dietetics, 111*(1):70-76. doi: 10.1016/j.jand.2012.09.018

Luce, RD & Raiffa H. (1989). Games and Decisions: Introduction and Critical Survey.

March, J.G. (1978). "Bounded rationality, ambiguity, and the engineering of choice." *The Bell Journal of Economics*. 587-608.

Moore, L.V., and Diez Roux, A.V. (2006). Associations of neighborhood characteristics with the location and type of food stores. *American Journal of Public Health, 96*(2), 325-331. doi: 10.2105/AJPH.2004.058040

Moore, L.V., Diez Roux, A.V., Nettleton, J.A., & Jacobs, D.R. (2008). Associations of the local food environment with diet quality—A comparison of assessments based on surveys and geographic information systems: The multi-ethnic study of atherosclerosis. *American Journal of Epidemiology, 167*(8), 917–924. doi: 10.1093/aje/kwm394

Morland, K., Diez Roux, A.V., & Wing, S. (2006). Supermarkets, other food stores, and obesity: The atherosclerosis risk in communities study. *American Journal of Preventive Medicine, 30*(4), 333-9.

Palmer, A., Smith, J., Haering, S. A., & McKenzie, S. (2007). *Understanding and addressing food security in Southwest Baltimore.* Unpublished manuscript, Center for a Livable Future. Retrieved from http://www.jhsph.edu/bin/o/u/OROSWreport2009-1-1.pdf

Pollard, S, Zachary, D, Wingert, K, Booker, S, Surkan, P. (2014). Family and community influences on diabetes-related dietary change in a low-income, urban neighborhood. *Diabetes Educator.* doi:10.1177/0145721714527520

Raja, S., Ma, C., & Yadav, P. (2008). Beyond food deserts: Measuring and mapping racial disparities in neighborhood food environments. *Journal of Planning Education and Research, 27*(4), 469-482. doi: 10.1177/0739456X08317461

Riegel, B., Dickson V., & Topaz M. (2013). Qualitative analysis of naturalistic decision making in adults with chronic heart failure, *Nursing Research, 62*(2): 91-98. doi: 10.1097/NNR.0b013e318276250c.

Rose, D.J. (2011). Captive audience? Strategies for acquiring food in two Detroit neighborhoods. Qualitative Health Research, *21*(5): 642-651. doi:10.1177/1049732310387159

Story, M., Kaphingst, K.M., Robinson-O'Brien, R., & Glanz K. (2008). Creating healthy food and eating environments: Policy and environmental approaches. *Annual Review of Public Health, 29.* 253-72. doi: 10.1146/annurev.publhealth.29.020907.090926

Sunstein, C.R., & Thaler, R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven: Yale University Press.

Treuhaft, S., & Karpyn, A. (2010). The grocery gap: Who has access to healthy food and why it matters. Philadelphia, PA: Policy Link, The Food Trust. Retrieved from http://www.policylink.org/grocerygap

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.

U.S. Department of Agriculture, Food and Nutrition Service, Office of Research and Analysis. 2014. SNAP Food Security In-Depth Interview Study. Kathryn Edin, Melody Boyd, James Mabli, Jim Ohls, Julie Worthington, Sara Greene, Nicholas Redel, Swetha Sridharan. Project Officer: Sarah Zapolsky, Alexandria, VA: March 2013.

U.S. Department of Agriculture, Food and Nutrition Service. (2013). Healthy Incentives Pilot. Retrieved from: http://www.fns.usda.gov/snap/hip/

VanLehn, K. (1996). Cognitive skill acquisition. *Annual review of psychology, 47*(1), 513-539.

Wingert, K, Zachary, D, Fox, M, Gittelsohn, J, Surkan, P. (2014). Child as change agent: The potential of children's influence to increase healthy food purchasing. *Appetite*, 81:330-336. doi: 10.1016/j.appet.2014.06.104

Woolf, S. H., & Braveman, P. (2011). Where health disparities begin: The role of social and economic determinants—and why current policies may make matters worse. *Health Affairs,* 30(10): 1852-1859. doi: 10.1377/hlthaff.2011.0685

Zachary, D., Palmer, A., & Surkan, P. (2012). Grocery purchasing in the food desert environment: Constraints outweigh preferences. Applied and Agricultural Economics Association Conference on "The Food environment: the effects of context on food choice," Boston, MA. Retrieved from: http://purl.umn.edu/123976

Zachary, D., Palmer, A., Beckham, S., & Surkan, P. (2013). A Framework for understanding grocery purchasing in a low-income urban community, *Qualitative Health Research, 23*(5): 665-678. doi: 10.1177/1049732313479451

Zachary, W. & Ryder, J. (1997). Decision Support Systems: Integrating Decision Aiding and Decision Training, in Helander, M, Landauer, T, Prabhu, P. *Handbook of Human-Computer Interaction*. Second, completely revised edition, Elsevier.

Zenk, S.N., Odoms-Young, A.M., Dallas, C., Hardy, E., Watkins, A., Hoskins-Wroten, J., Holland, L. (2011). "You have to hunt for the fruits, the vegetables": Environmental barriers and adaptive strategies to acquire food in a low-income African American neighborhood. Health Education & Behavior, 38(3): 282-292. doi:10.1177/1090198110372877

Middlesex University London

Distributed Sensemaking: A Case Study of Military Analysis

Simon Attfield[a], Bob Fields[a], Ashley Wheat[a], Rob Hutton[b], Jim Nixon[c] , Andrew Leggatt[b] and Hannah Blackford[b].

[a] *Interaction Design Centre, Middlesex University, UK.*

[b] *Advanced Technology Centre, BAE Systems, UK*
(now with TriMetis Ltd., UK).

[c] Centre for Safety and Accident Investigation, Cranfield University, UK.

The Problem:

Sensemaking:

- comprehension (Klein et al., 2007)
- finding meaning from information (Weick, 1995)
- constructing 'pictures' of the world inferred from data (and background beliefs)

Sensemaking and Distributed Cognition:

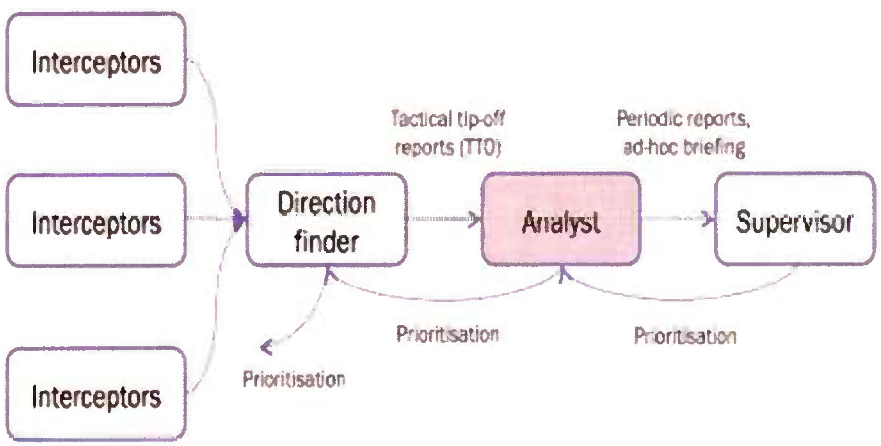
- sensemaking not just 'in the head' but in the world
- complete account of cognition needs to consider how it distributes across materials, time and people (Hutchins, 1995)
- external representations:
  - many forms (lists, maps, charts, pictures, reference information)
  - embodied in media/artefacts with different properties and affordances
- current theories tend not to engage in depth with how representational artefacts support, mediate and enhance cognition in sensemaking

Overall aim

- to contribute to better understanding of sensemaking as ‘distributed sensemaking’
- look at how external representations support reasoning during sensemaking
- we consider a sensemaking task (military signals intelligence exercise) which involves the use of external representational artefacts.
- attend to how artefacts *support*, *mediate* and *enhance* cognition.

Observation

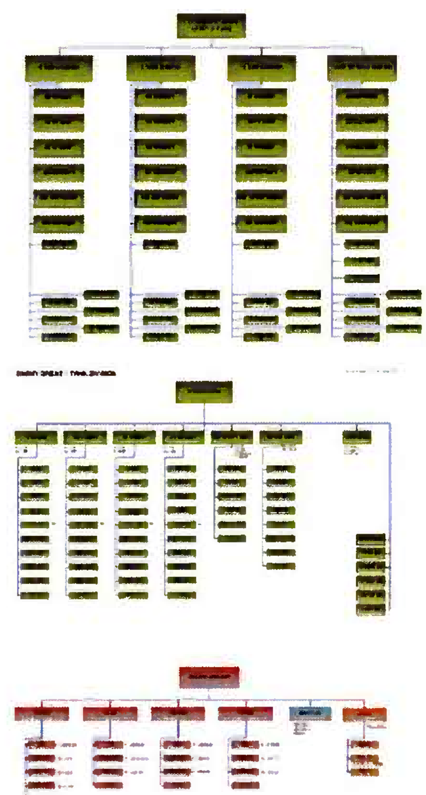
- SIGINT training exercise
- ex-military personnel
- screen recording, video recording, audio recording periodic debriefs

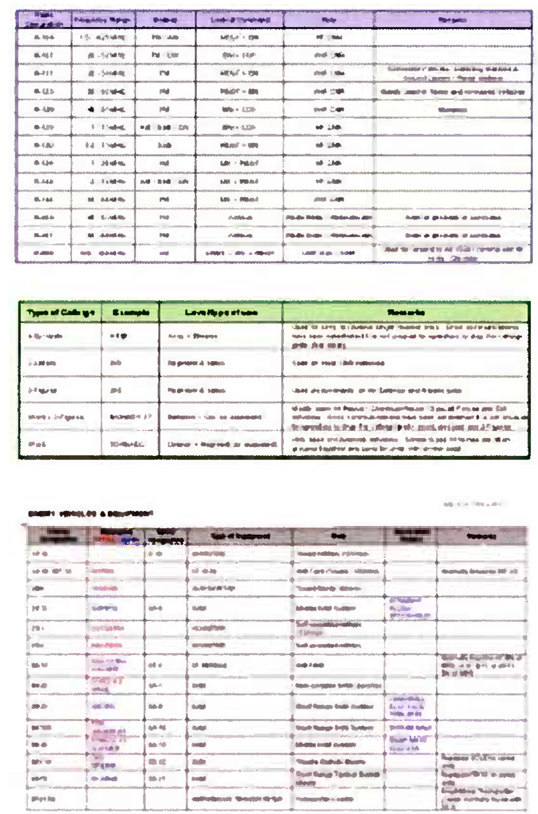


Case-study

Printed materials

ORBAT





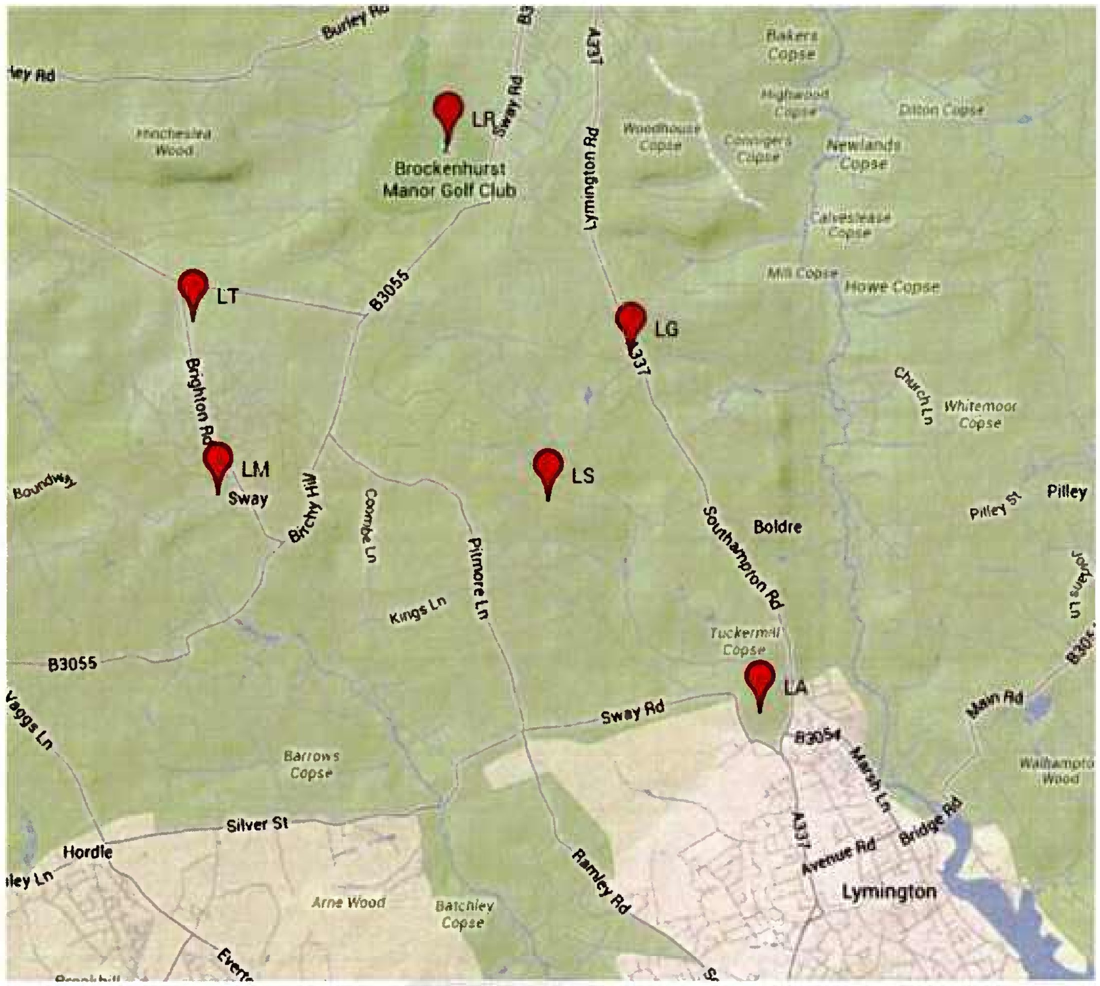
tables

Findings

First step – plot entities on map

TTO

lat/longs,
callsigns



Findings

TTO

3.55Mz
FM

Using the Radio Equipment Table

Radio Designation	Frequency Range	Mode(s)	Level of Command	Role	Remarks
P-404	1.25 - 4.5 MHz	FM / AM	REGT > BN	HF CNR	
~~P-407~~	~~20 - 52 MHz~~	~~FM / CW~~	~~BN > COY~~	~~VHF CNR~~	
~~P-411~~	~~20 - 48 MHz~~	~~FM~~	~~REGT > BN~~	~~VHF CNR~~	~~Command Vehicles, Listening Stations & Ground Launch / Radar stations~~
~~P-420~~	~~20 - 52 MHz~~	~~FM~~	~~REGT > BN~~	~~VHF CNR~~	~~Mainly used in Tanks and Armoured Vehicles~~
~~P-426~~	~~30 - 51 MHz~~	~~FM~~	~~BN > COY~~	~~VHF CNR~~	~~Manpack~~
~~P-429~~	~~1 - 10 MHz~~	~~AM / SSB / CW~~	~~BN > COY~~	~~HF CNR~~	
~~P-430~~	~~1.5 - 13.5 MHz~~	~~SSB~~	~~REGT > BN~~	~~HF CNR~~	
P-434	1 - 30 MHz	AM	DIV > REGT	HF CNR	
~~P-443~~	~~2 - 20 MHz~~	~~AM / SSB / CW~~	~~DIV > REGT~~	~~HF CNR~~	
~~P-445~~	~~32 - 62 MHz~~	~~FM~~	~~DIV > REGT~~	~~VHF CNR~~	
~~P-603~~	~~40 - 51 MHz~~	~~FM~~	~~Various~~	~~Radio Relay / Rebroadcast~~	~~Seen at all levels of command~~
~~P-707~~	~~52 - 60 MHz~~	~~FM~~	~~Various~~	~~Radio Relay / Rebroadcast~~	~~Seen at all levels of command~~
~~P-709~~	~~100 - 150 MHz~~	~~AM~~	~~ARMY > DIV > REGT~~	~~UHF G2A / CNR~~	~~Used for Ground to Air (G2A) comms and for Army - Div links~~

Findings

Using the Callsign Table

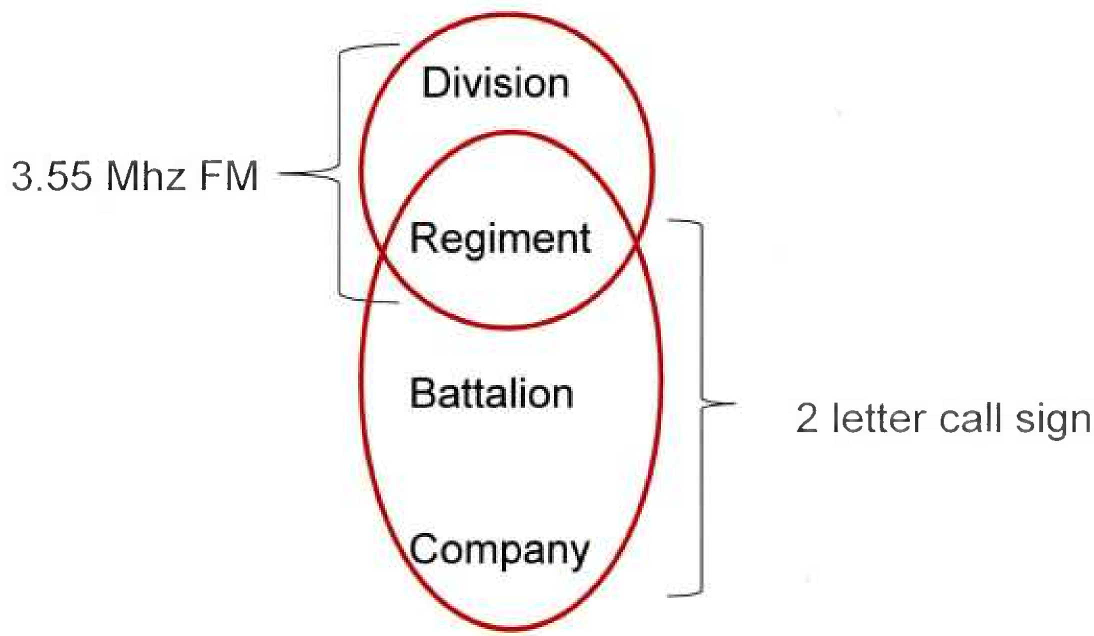
TTO

Callsigns
LO
LR
LM
LP
LQ

Type of Callsign	Example	Level/type of use
~~4-Symbols~~	~~S21A~~	~~Army > Division~~
2-Letters	NE	Regiment & below
~~3-Figures~~	~~245~~	~~Regiment & below~~
~~Word + 2-Figures~~	~~LION 17~~	~~Battalion > Coy (or equivalent)~~
~~Word~~	~~TORPEDO~~	~~Division > Regiment (or equivalent)~~

Findings

1 Combined Arms Army



Conceptualising Resources via Toulmin's model of argument

Representations have different roles

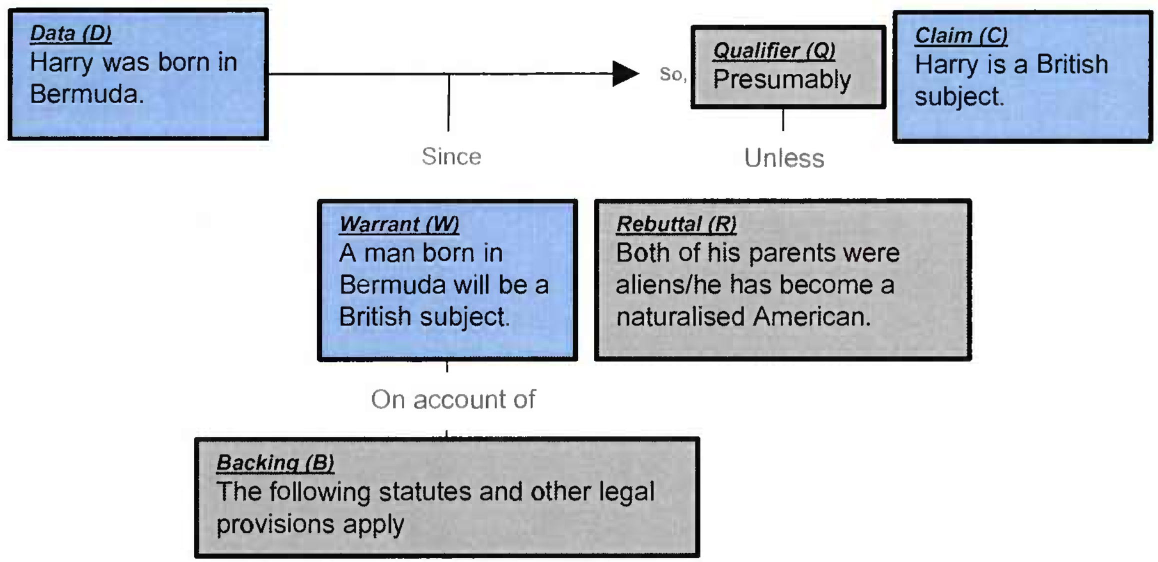


Table 1. Cues (left), enabled information to be inferred (right) given mediating representations (middle).

Data	Warrant representation	Claim
Frequency	Radio Equipment table	Level of command
Call sign	Radio Procedures – Call signs table	
Type of encryption	ROM Encryption Systems Table	
Level of command, code words, call signs and message extracts	ORBAT	Unit identity
	Background knowledge	Action

Representations and artefact affordances supported deduction by *modus tolens*

‘Warrant’ information (associative, rule-like) is often tacit.

Given desired computation and material properties of representational artefact, analysts constructed strategies with particular properties.

Physical level

- Testing values against statements of frequency range
- Annotating on outcome
- Comparing unannotated entries

Computational level

- Maintaining a set of working hypotheses
  - ‘Logical Competitor Set’ (Feltovich et al 1984)
  - ‘Differential’ (Josephson and Josephson, 1996)
- Inference in sensemaking noted as often abductive (Klein et al., 2007)
  - But process was a *modus tolens* inference i.e. deductive.
  - i.e. rigorous and even ‘normative’

We assume through ‘seeing’ the strategy as a possibility (given artefact affordances) the computation becomes a possibility.

modus tollens

if A then B

not B

therefore not A

if P407 then 28-50 Mhz

not 28-50Mhz

therefore not P407

And finally,

- Sensemaking can be analysed as distributed cognition.
- Highlights different roles played by different representational artefacts.
- Distributed sensemaking is seen as a process of transforming and propagating representational state in order to make interpretations, consider ‘competitor sets’, alternative hypotheses or ‘frames’, and develop a rich and reliable situation picture.
- We believe that this approach contributes towards a more informed framework (language) for exploring and considering sensemaking design requirements.

ACKNOWLEDGMENTS

This work was funded by the UK Defence Science and Technology Laboratory (Dstl) via the UK Defence Human Capability Science and Technology programme. The training exercise was devised and run by MASS Consultants Ltd.

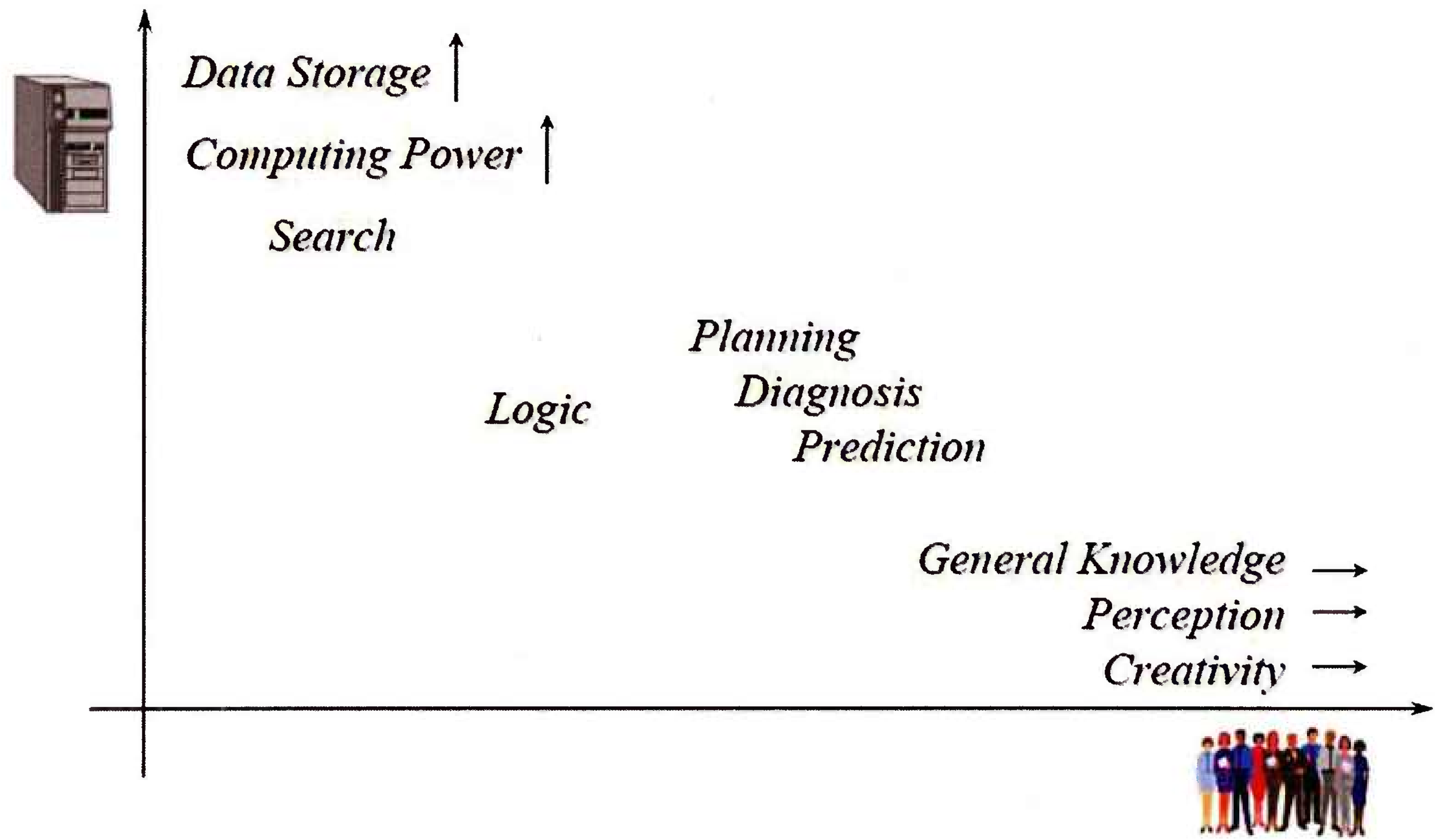
Visual Analytics - definition

The science of analytical reasoning facilitated by interactive visual interfaces.

Analytical Reasoning
Visual Representations and interaction technologies
Data Representations and transformations
Production, presentation and dissemination

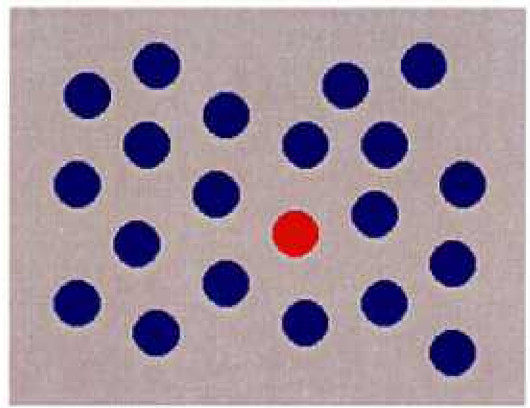
(Thomas and Cook, 2006)

Comparing the abilities of people and computers



(Keim, 2007)

The power of pre-attentive processing



	A	B	C
1	**Animal**	**Body Weight (kg)**	**Brain Weight (kg)**
2	Mountain beaver	1.35	465
3	Cow	465	423
4	Grey wolf	36.33	119.5
5	Goat	27.66	115
6	Guinea pig	1.04	5.5
7	Dipliodocus	11700	50
8	Asian elephant	2547	4603
9	Donkey	187.1	419
10	Horse	521	655
11	Potar monkey	10	115
12	Cat	3.3	25.6
13	Giraffe	529	680
14	Gorilla	207	406
15	Human	62	1320
16	African elephant	6654	5712
17	Triceratops	9400	70
18	Rhesus monkey	6.8	179
19	Kangaroo	35	56
20	Golden hamster	0.12	1
21	Mouse	0.023	0.4
22	Rabbit	2.5	12.1
23	Sheep	55.5	175
24	Jaguar	100	157
25	Chimpanzee	52.16	440
26	Brachiosaurus	87000	154.5
27	Mole	0.122	3
28	Pig	192	180

(Rousseuw & Leroy, 1987)

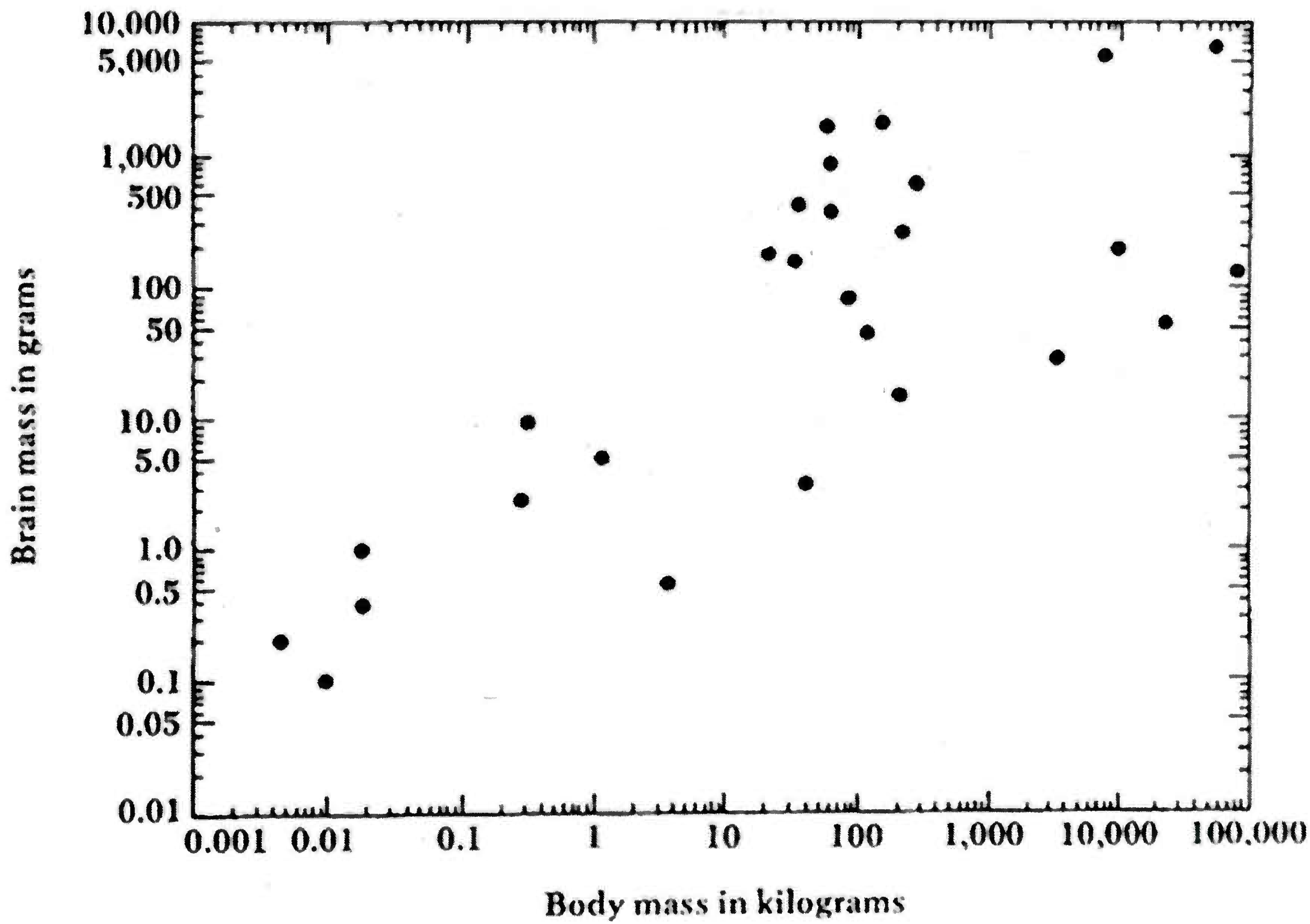
	A	B	C	D
1	Animal	Brain Weight (kg)	Body Weight (kg)	
2	Mountain beav	1.35	465	
3	Cow	465	423	
4	Grey wolf	36.33	119.5	
5	Goat	27.66	115	
6	Guinea pig	1.04	5.5	
7	Dipliodocus	11700	50	
8	Asian elephant	2547	4603	
9	Donkey	187.1	419	
10	Horse	521	655	
11	Potar monkey	10	115	
12	Cat	3.3	25.6	
13	Giraffe	529	680	
14	Gorilla	207	406	
15	Human	62	1320	
16	African elephar	6654	5712	
17	Triceratops	9400	70	
18	Rhesus monke	6.8	179	
19	Kangaroo	35	56	
20	Golden hamste	0.12	1	
21	Mouse	0.023	0.4	

Which animal weighs the least/most?

Is there a relationship between brain weight and body weight?

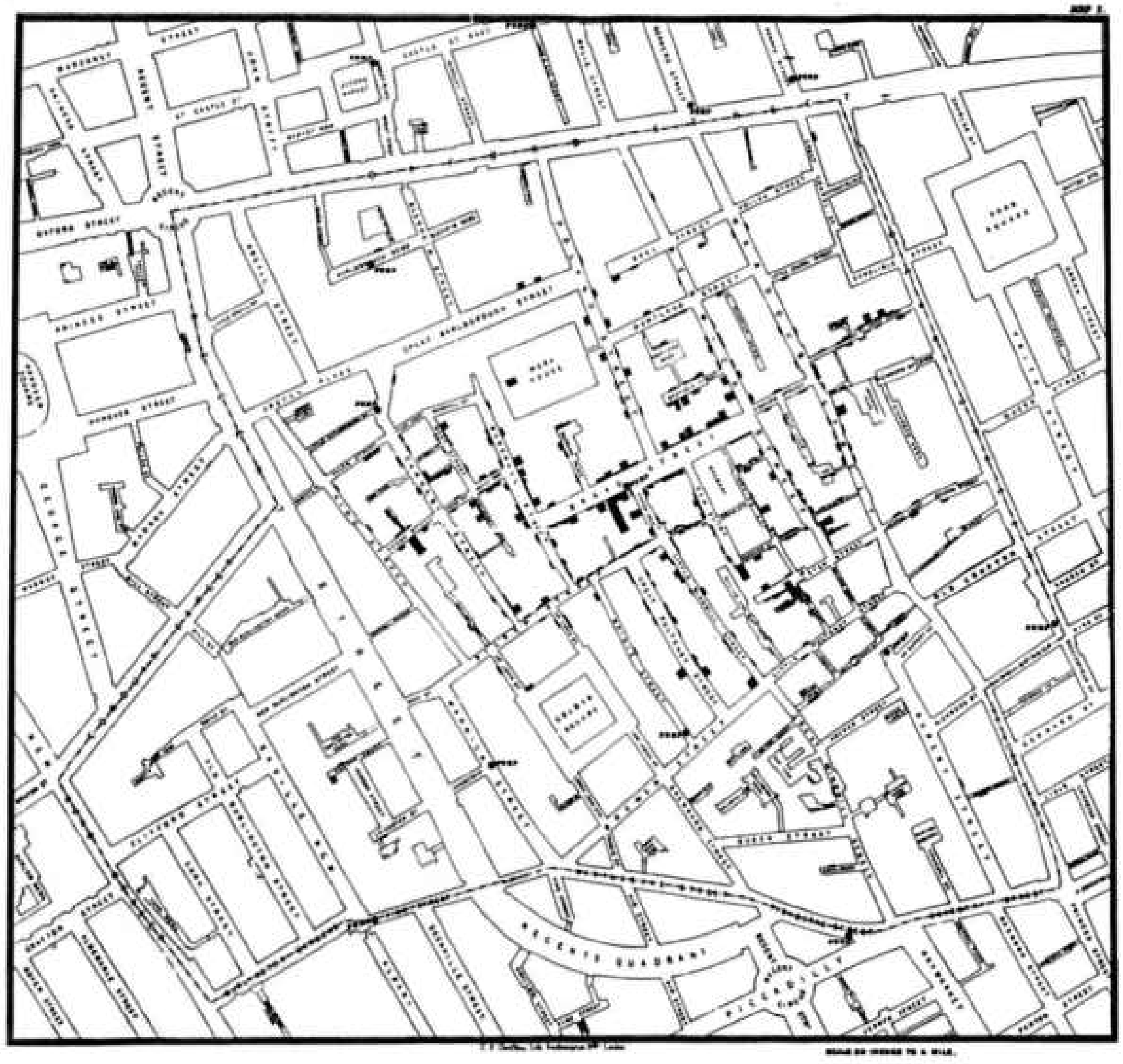
If so, are there any outliers?

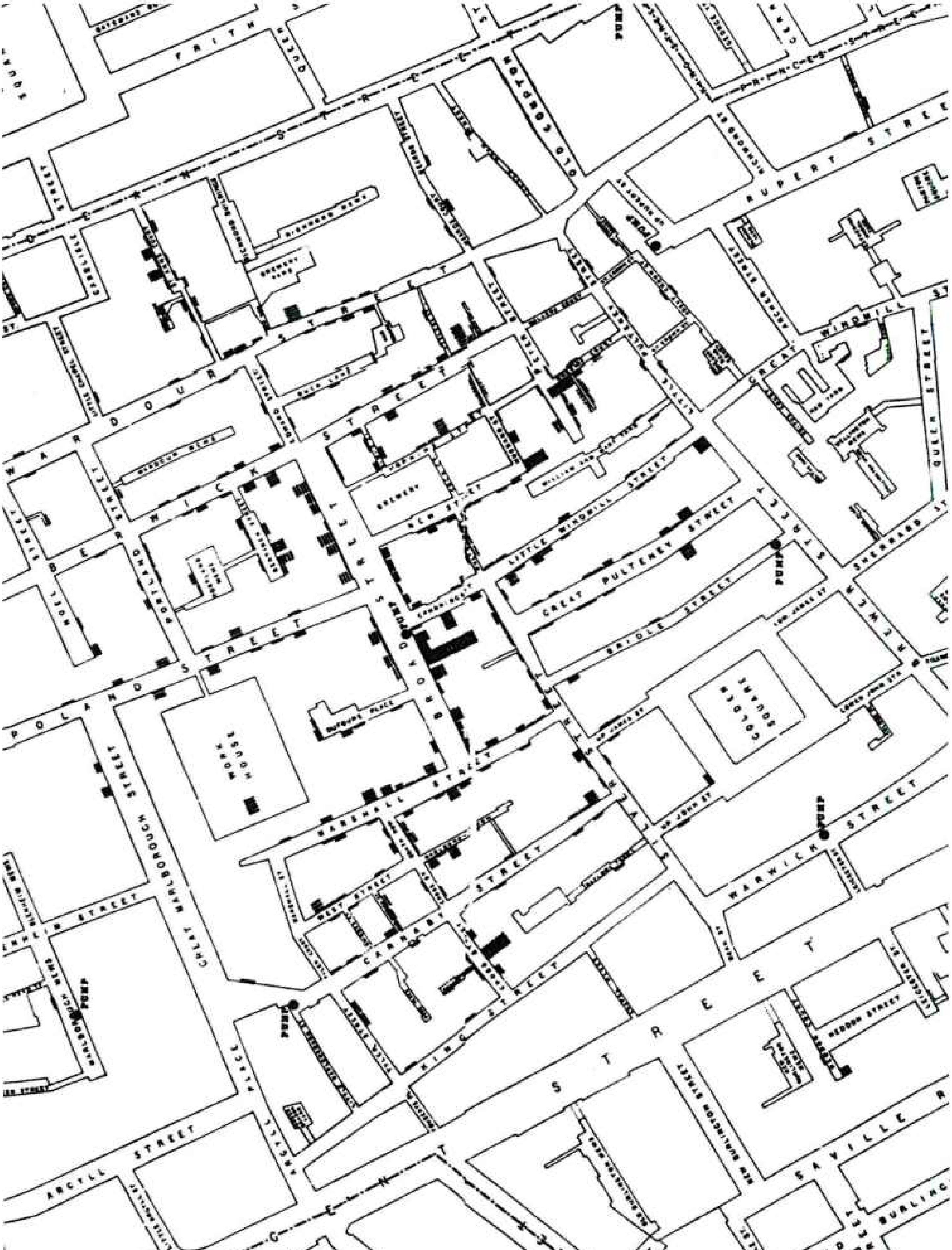
Rousseeuw, P.J. & Leroy, A.M. (1987) Robust Regression and Outlier Detection. Wiley, p. 57



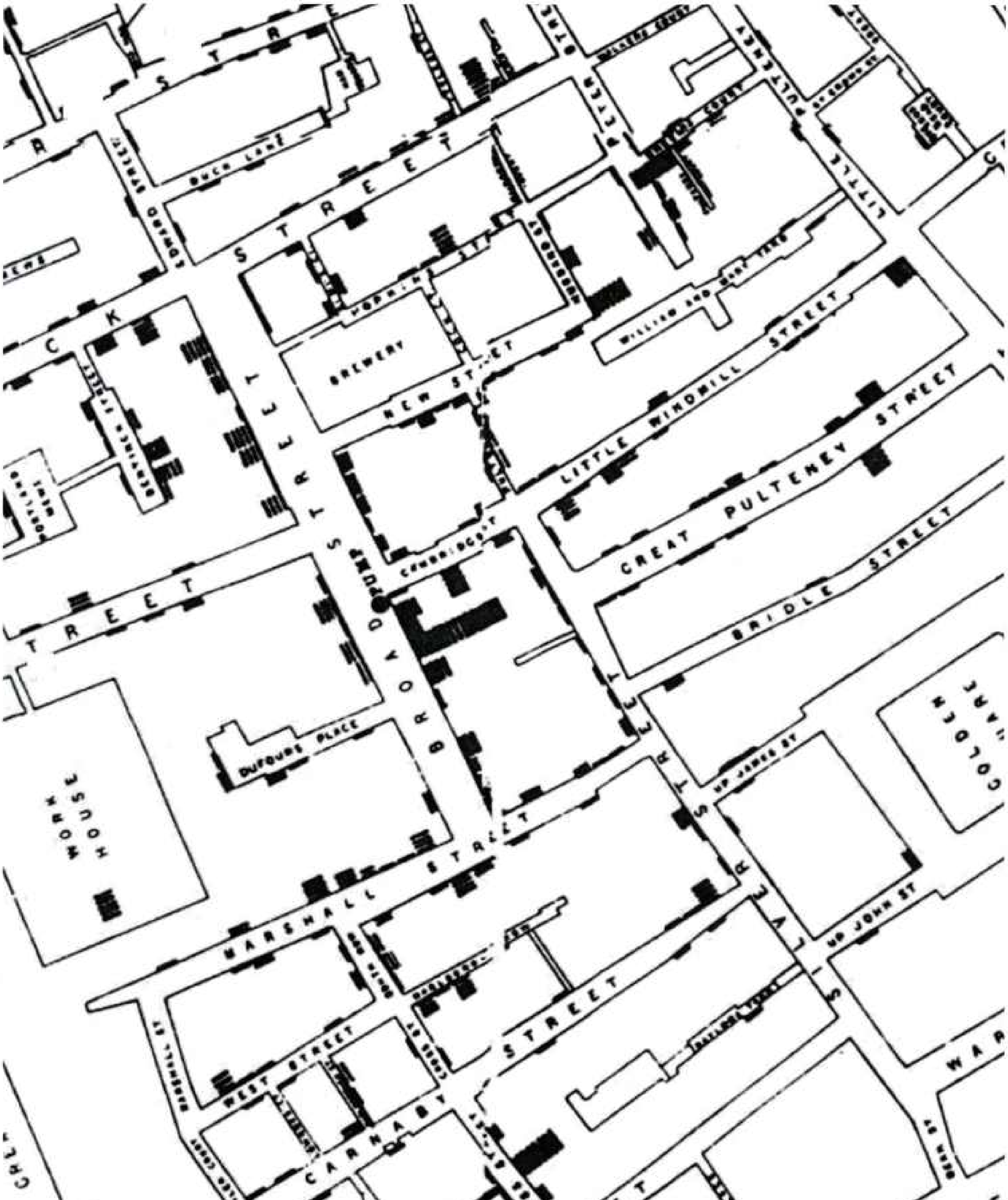
Brain mass in grams
10,000
5,000
1,000
500
100
50
10.0
5.0
1.0
0.5
0.1
0.05
0.01
0.001
0.01
0.1
1
10
100
1,000
10,000
100,000
Body mass in kilograms

John Snow's Cholera map



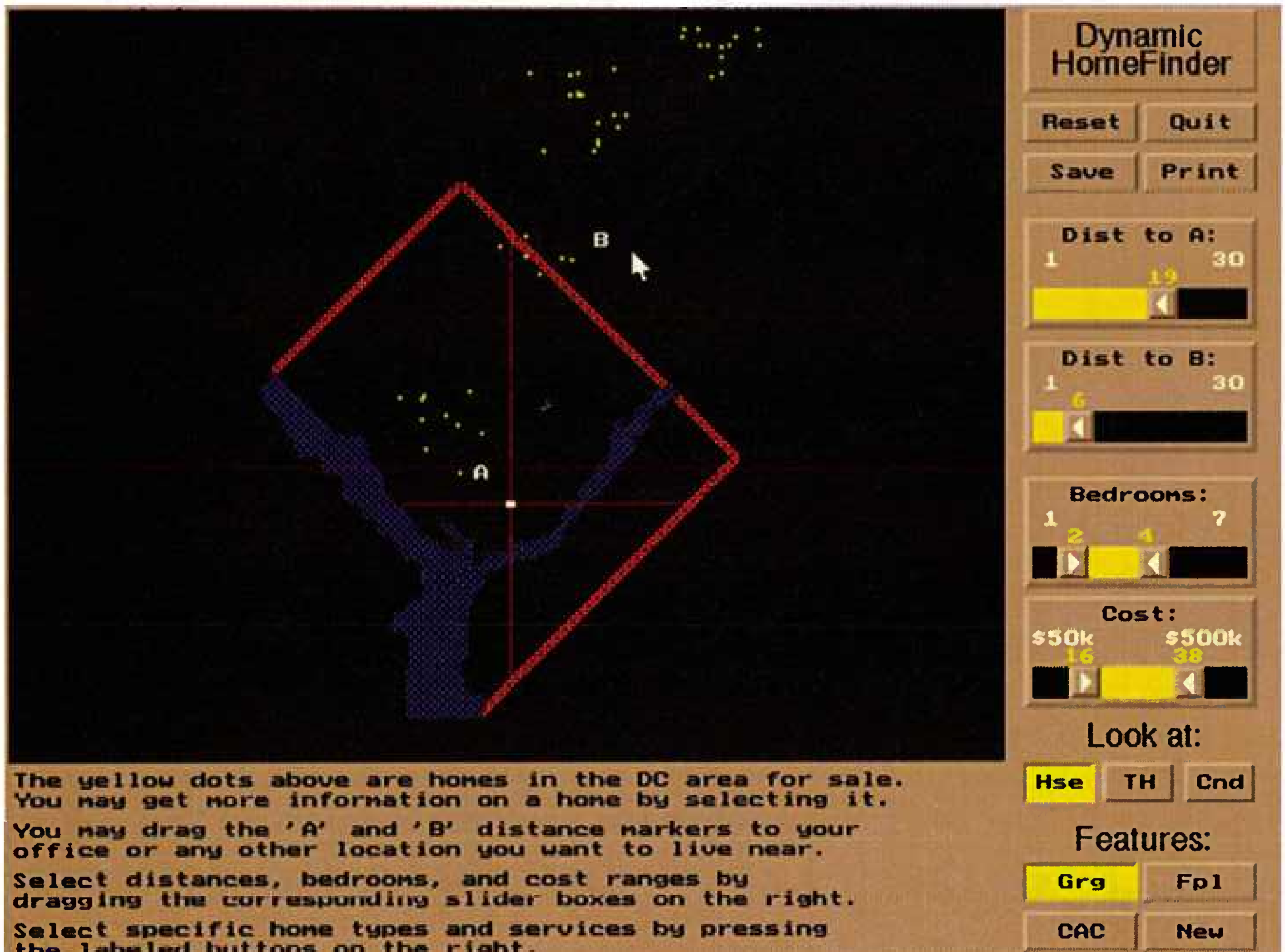


FRITH
QUEEN
OLD COMPTON
RUPERT STREET
CARLISLE STREET
BREWERY YARD
WARDOUR
WARDOUR MEWS
BERWICK
NOEL STREET
PORTLAND STREET
POLAND STREET
BROAD STREET
BREWERY
CAMBRIDGE ST
DUFOURS PLACE
WORK HOUSE
LITTLE WINDMILL STREET
GREAT PULTENEY STREET
BRIDLE STREET
GOLDEN SQUARE
GREAT WINDMILL
QUEEN STREET
SHERRARD ST
MARSHALL STREET
GREAT MARLBOROUGH STREET
CARNABY
KING STREET
WARWICK STREET
HEDDON STREET
NEW BURLINGTON STREET
ARGYLL STREET
SAVILLE R
PUMP
PUMP
PUMP
PUMP
PUMP
PUMP

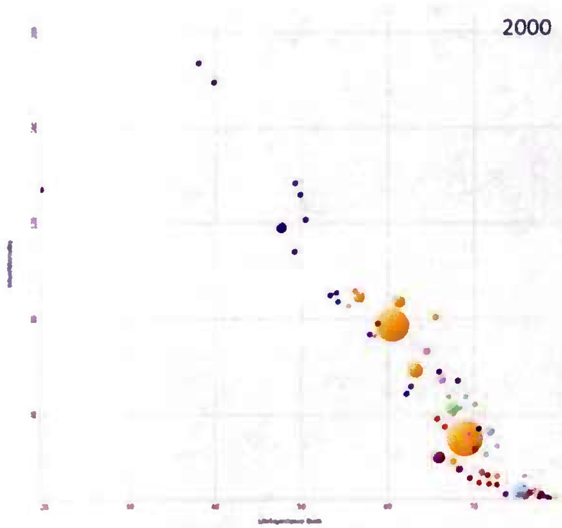


BROAD STREET
PUMP
CAMBRIDGE ST
NEW STREET
BREWERY
LITTLE WINDMILL STREET
GREAT PULTENEY STREET
BRIDLE STREET
MARSHALL STREET
DUFOURS PLACE
WORK HOUSE
CARNABY STREET
WEST STREET
MARSHALL ST
DUCK LANE
EDWARD STREET
BERWICK STREET
PETER STREET
LITTLE PULTENEY
UP JOHN ST
UP JAMES ST
GOLDEN SQUARE
KING STREET
PUMP

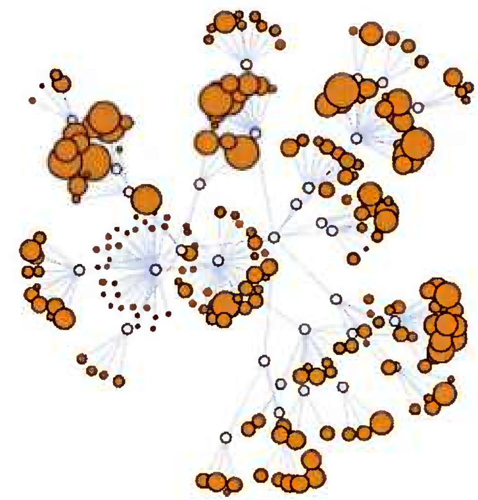
Interactive visualisation
Christopher Williamson's HomeFinder



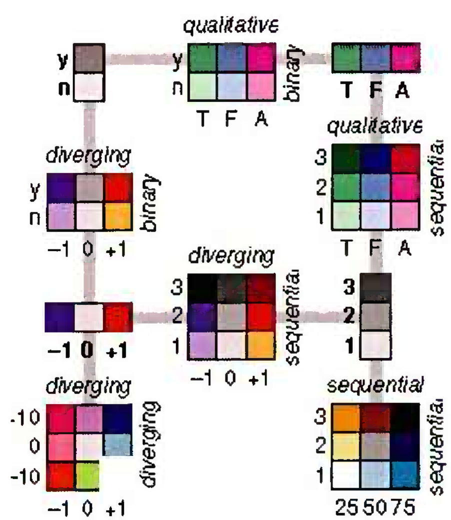
Common Visualisation Techniques



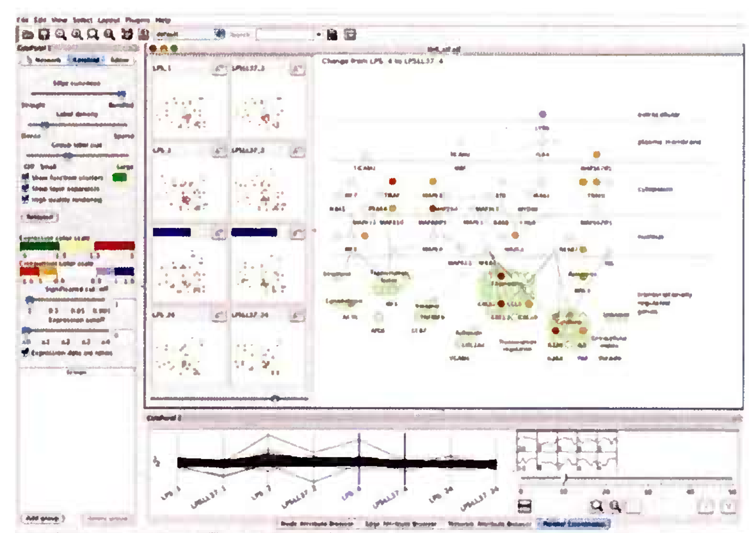
Charts



Trees and graphs



Colours



Putting them together

- Negotiate the requirement – what are the ‘givens’ and the ‘changeables’.
- Speak many tongues and be agile (go round the roundabout)
- Conceptualise at many levels of abstraction considering both need and opportunity

Dealing with the "Un-ness" of Events

- Emerging Challenges can be characterised by their 'un-ness':
  - Unexpected: indicating problems in the ways in which events can be predicted, primarily in terms of *when* they will occur but also in terms of *where* they will occur.
  - Unprecedented: indicating problems in the knowledge-base relating to such events, and the prior experiences that could be brought to bear in dealing with these events.
  - Unmanageable: indicating problems in defining and resourcing response to the events.

Hewitt, K. (1983) *Interpretations of Calamity: From the Viewpoint of Human Ecology*, London: Allen and Unwin.

UNIVERSITY OF BIRMINGHAM

Working together...

UNIVERSITY OF BIRMINGHAM

...working apart...

- ☐ Challenges of distributed working
- ☐ Challenges of interoperability
- ☐ Challenges of ‘common ground’

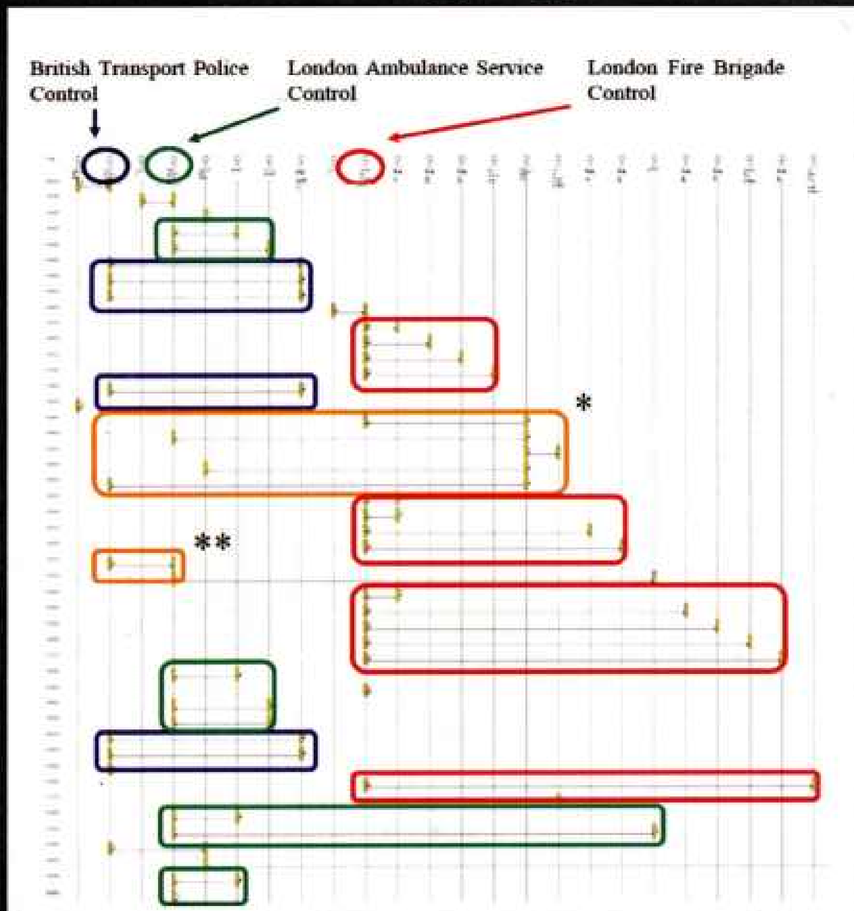
UNIVERSITY OF BIRMINGHAM

Challenge of Sharing information

- Data and Information Semantics
  - Common languages
  - Common Operating Pictures
  - Cross-Agency common ground?
- Data Quality, Quantity and Timeliness
  - Socio-Technical Interoperability
- Data and Information Relevance
  - Relevant for whom?
  - Relevant for what?

UNIVERSITY OF BIRMINGHAM

...working apart....



- Stove-piping across Agencies
- How does this change when Agencies are pursuing different goals or dealing with different 'situations'?

UNIVERSITY OF BIRMINGHAM

Semantic and Pragmatic Sensemaking

- Semantic: Knowing what to say
- Pragmatic: Knowing who to say it to and when to say it

Discussion Points

- □ NDM in multi-agency operations
- □ NDM in situations of "un-ness"
- □ Gaining access to critical people, places and situations
- □ Capturing activity data in situ

UNIVERSITY OF BIRMINGHAM

Cyber-Physical Threats in the Food Industry: Toward Real-Time Anticipation, Detection, Response, Adaptation, and Recovery

Jeffrey M. Bradshaw, Larry Bunch
{jbradshaw, lbunch}@ihmc.us
www.ihmc.us/groups/jbradshaw/

ihmc

ihmc

Examples of IHMC Work in Cyber Sensemaking

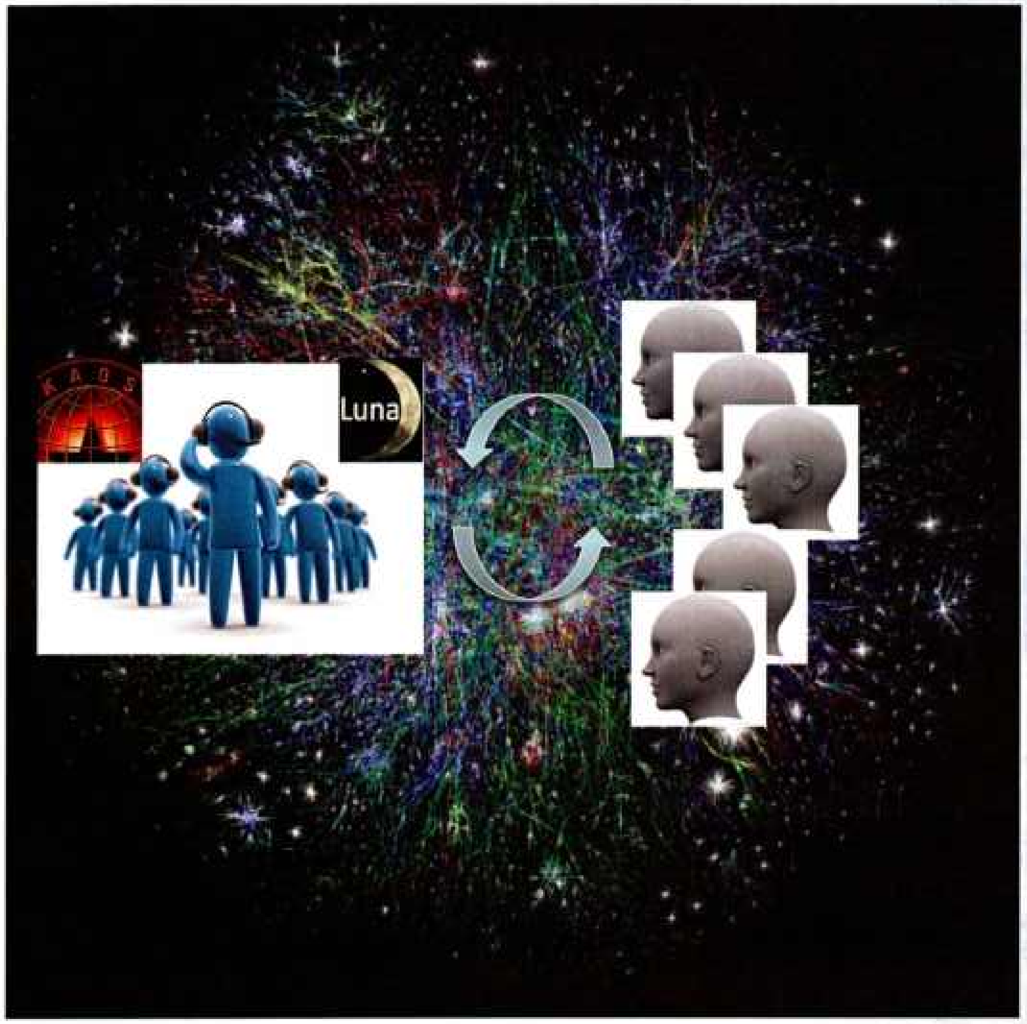
- Pioneer in Cognitive Task Analysis
- Funded by DoD to develop cyber defense technologies that leverage joint human-machine teamwork
- Sponsored by MIT Lincoln Lab to develop CCER standards for semantic representation of reusable cyber range models; use in cyber exercises
- KAoS ontology selected as the basis for NSA Digital Policy Management initiative
- ONR: Red Team in a Box (SoarTech)
- Capstone Training Project (TCS/Center for Information Dominance)
- ITADS (Intelligent Tutoring - TCS/Center for Information Dominance)
- Explorations with National Center for Food Protection and Defense (NCFPD)

Foundational Principles

- Visualization tasks are situated in larger cognitive and social contexts of sensemaking
  - Focus on what analysts need to know, not what is easy to show
- Sensitivity to visual aesthetics needs to be coupled with an understanding of human perceptual and cognitive processes
- Normative performance models help focus people on what is important or anomalous
- Policy-governed software agents amplify visual salience of events of interest to make them stand out from the "noise"

ihmc

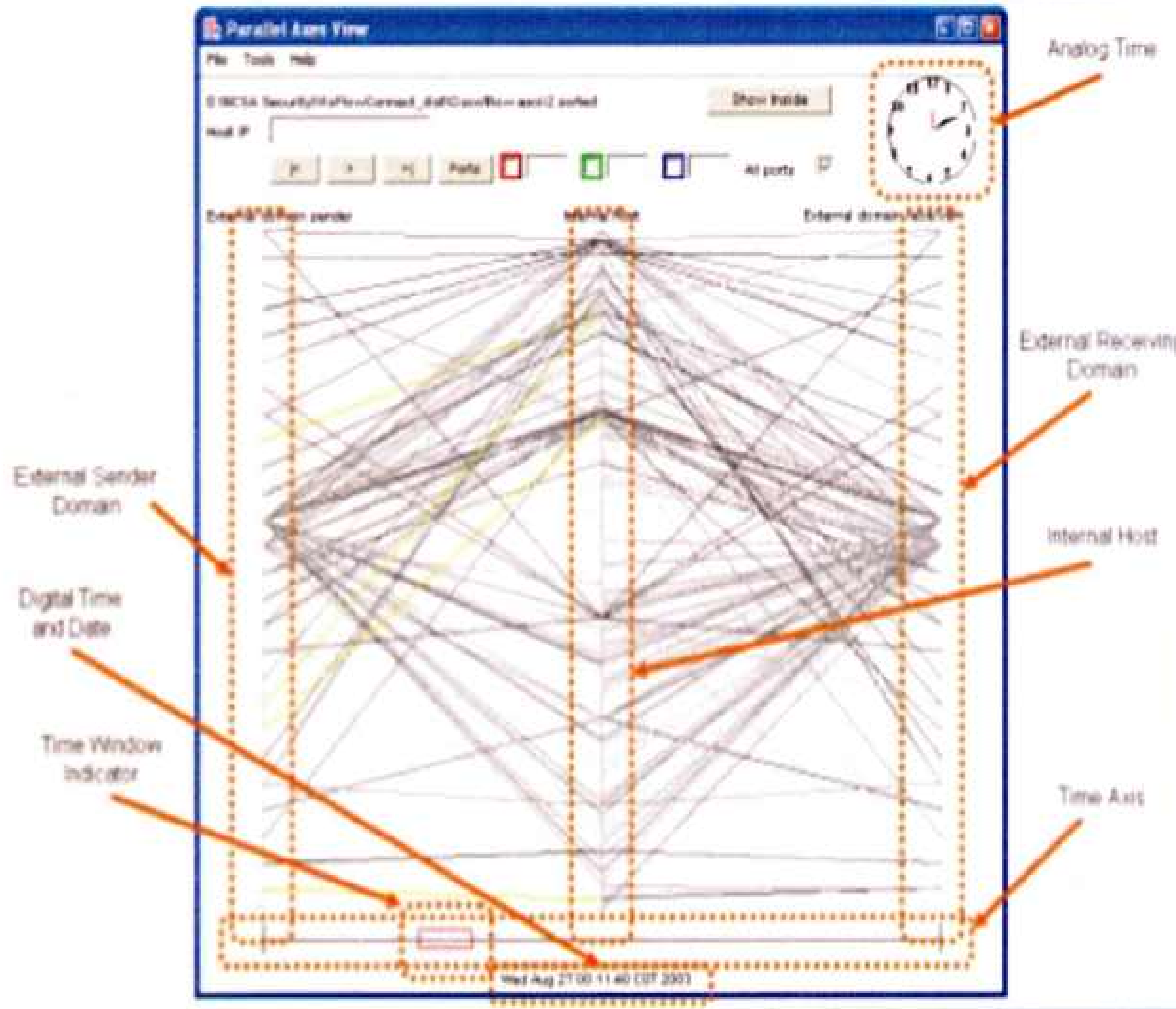
Key Capabilities for Distributed Sensemaking



- *OZ-Inspired Visualizations* based on principles of human perception and cognition, and designed around what people need to know in a given context
- *Luna Software Agents* based on an understanding of joint activity and biologically-inspired learning, and designed to assist team members with a variety of tasks
- *KAoS: An Adaptive, Policy-Governed System Architecture* designed to ensure directability and resilience

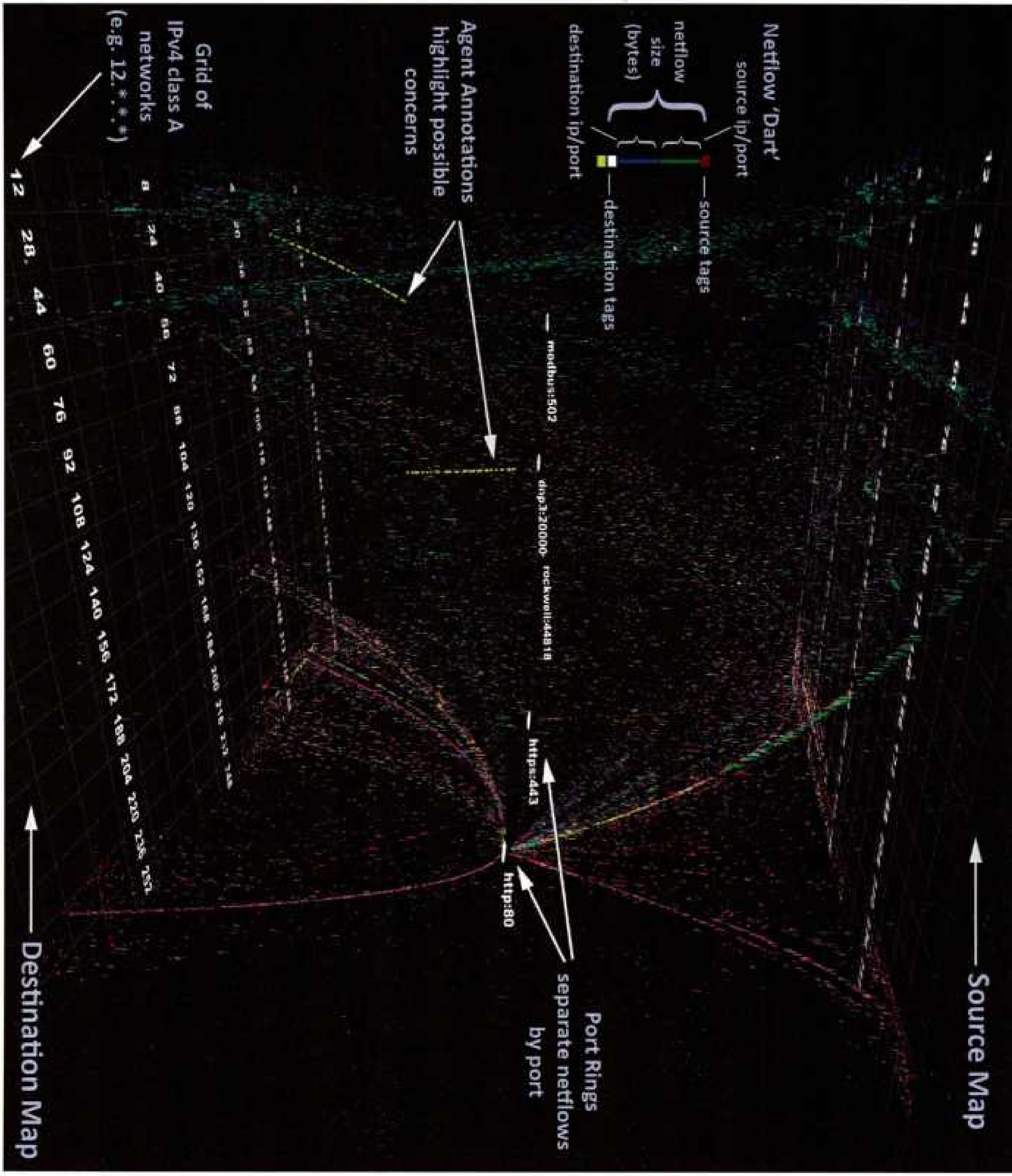
ihmc

VisFlowConnect Parallel Coordinates Display

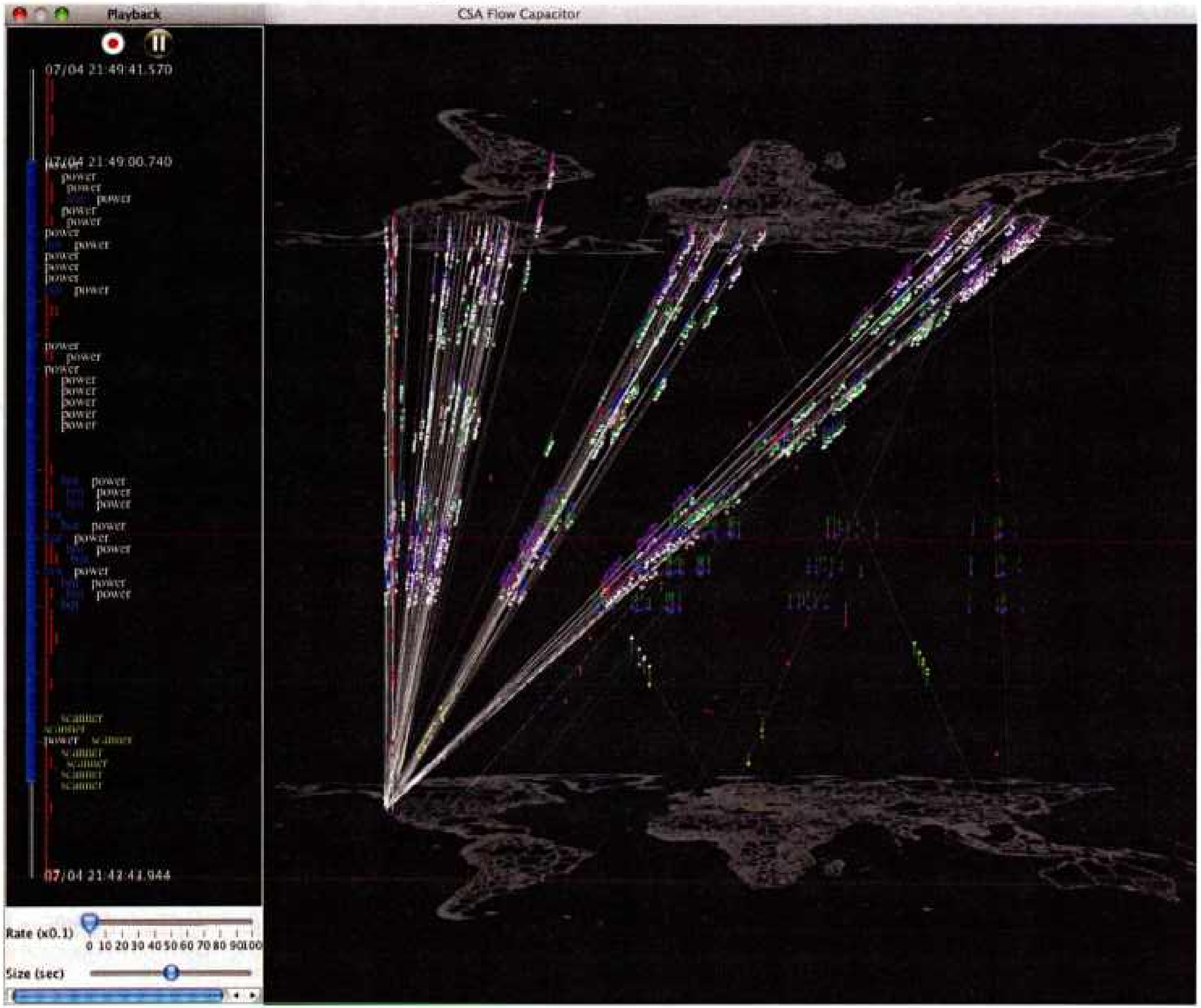


ihmc

Network Observatory



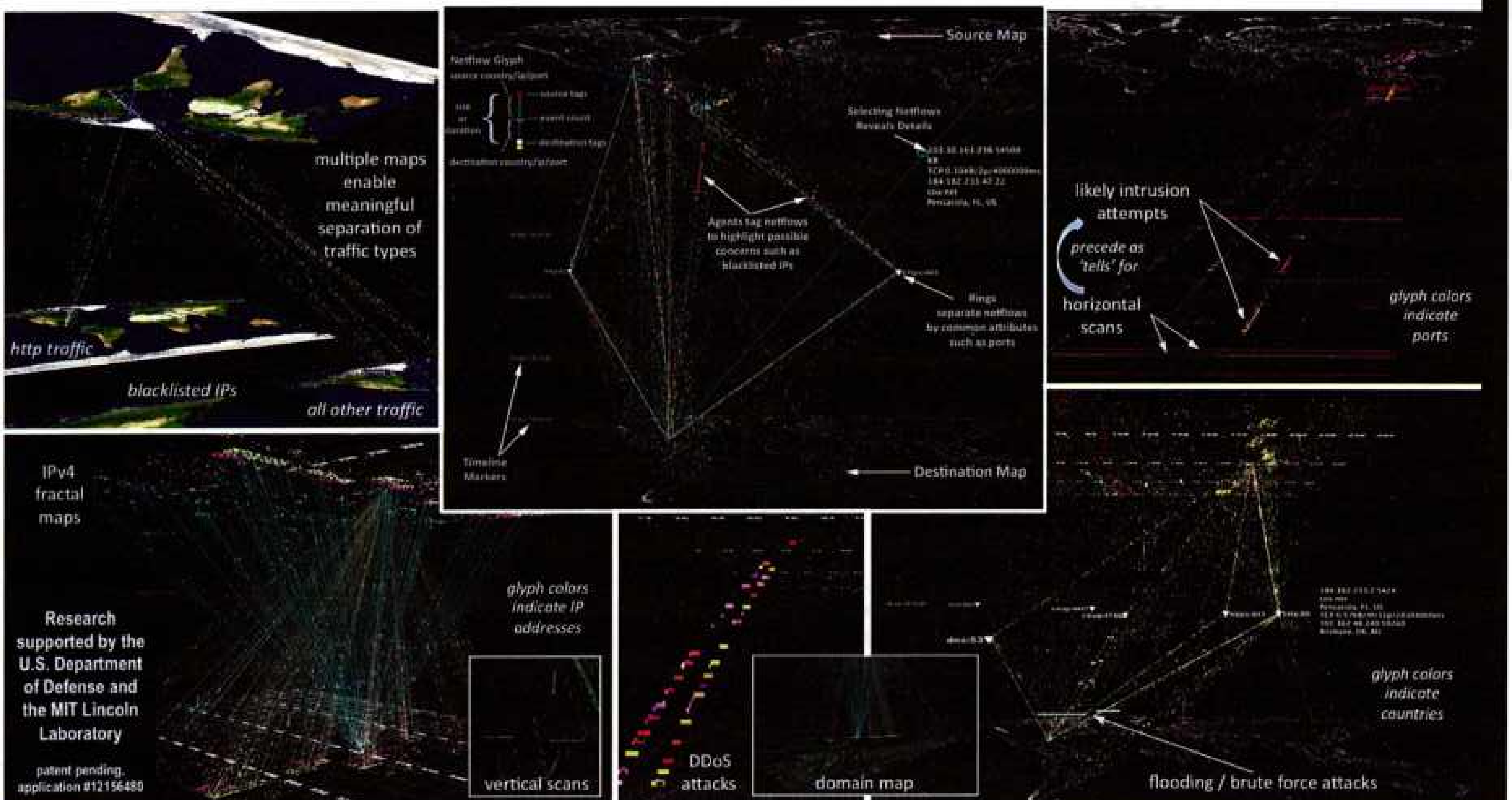
Anatomy of a DDOS Attack



ihmc

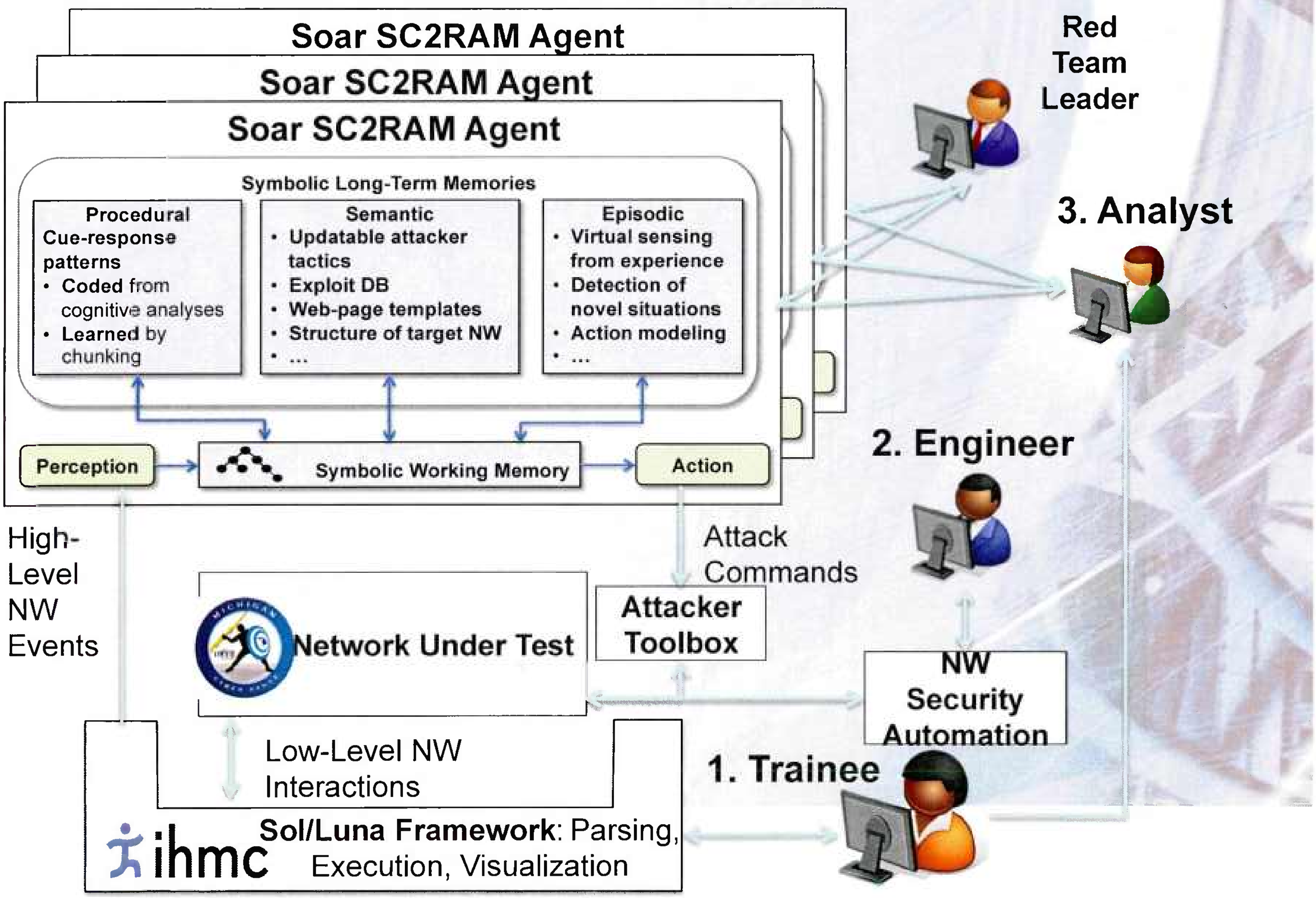
The Sol Netflow Observatory

An Agent-Based Interactive Cyber Defense Framework

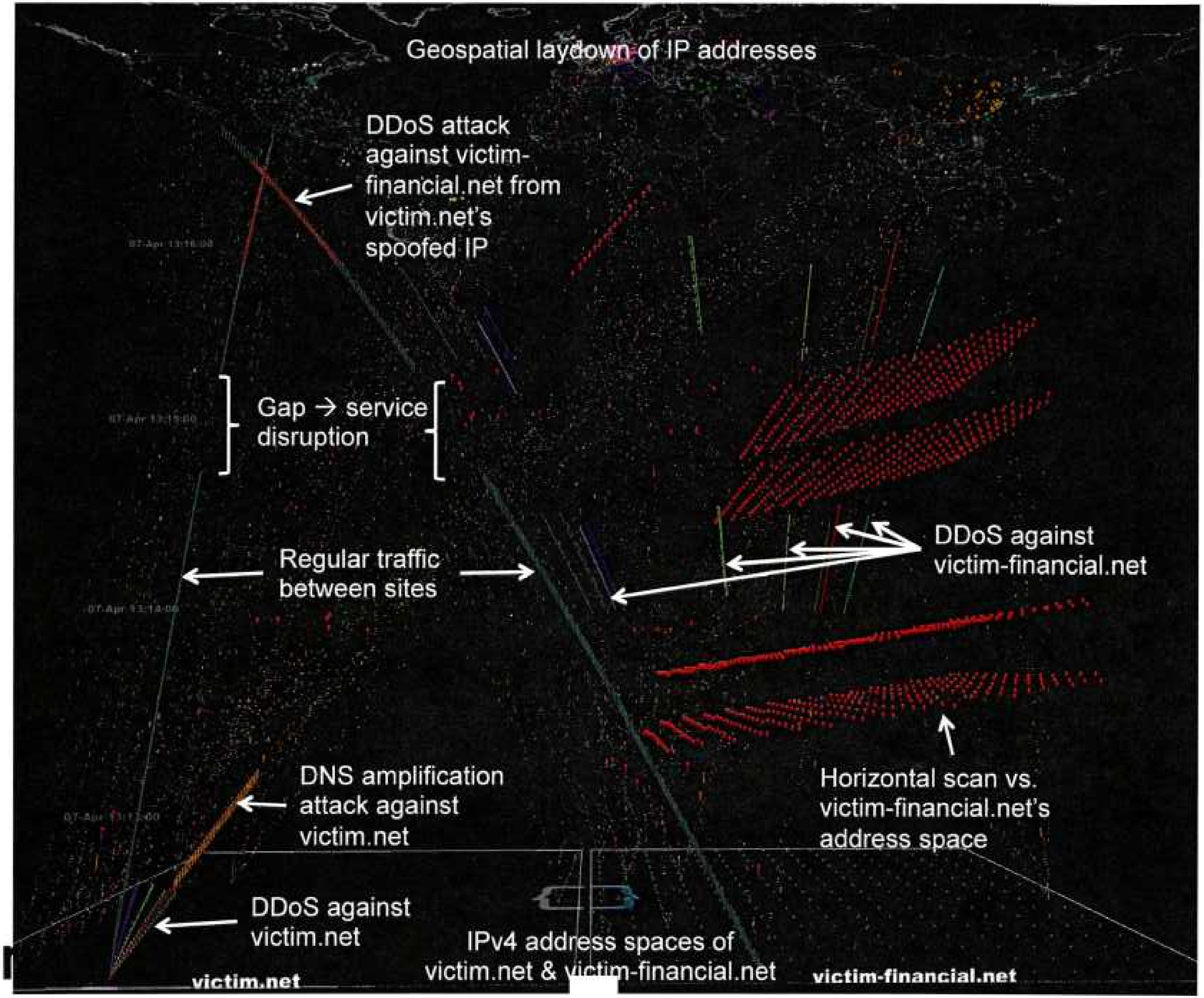


ihmc

ONR Cyber "Red-Team-In-a-Box"



Phase I Scenario

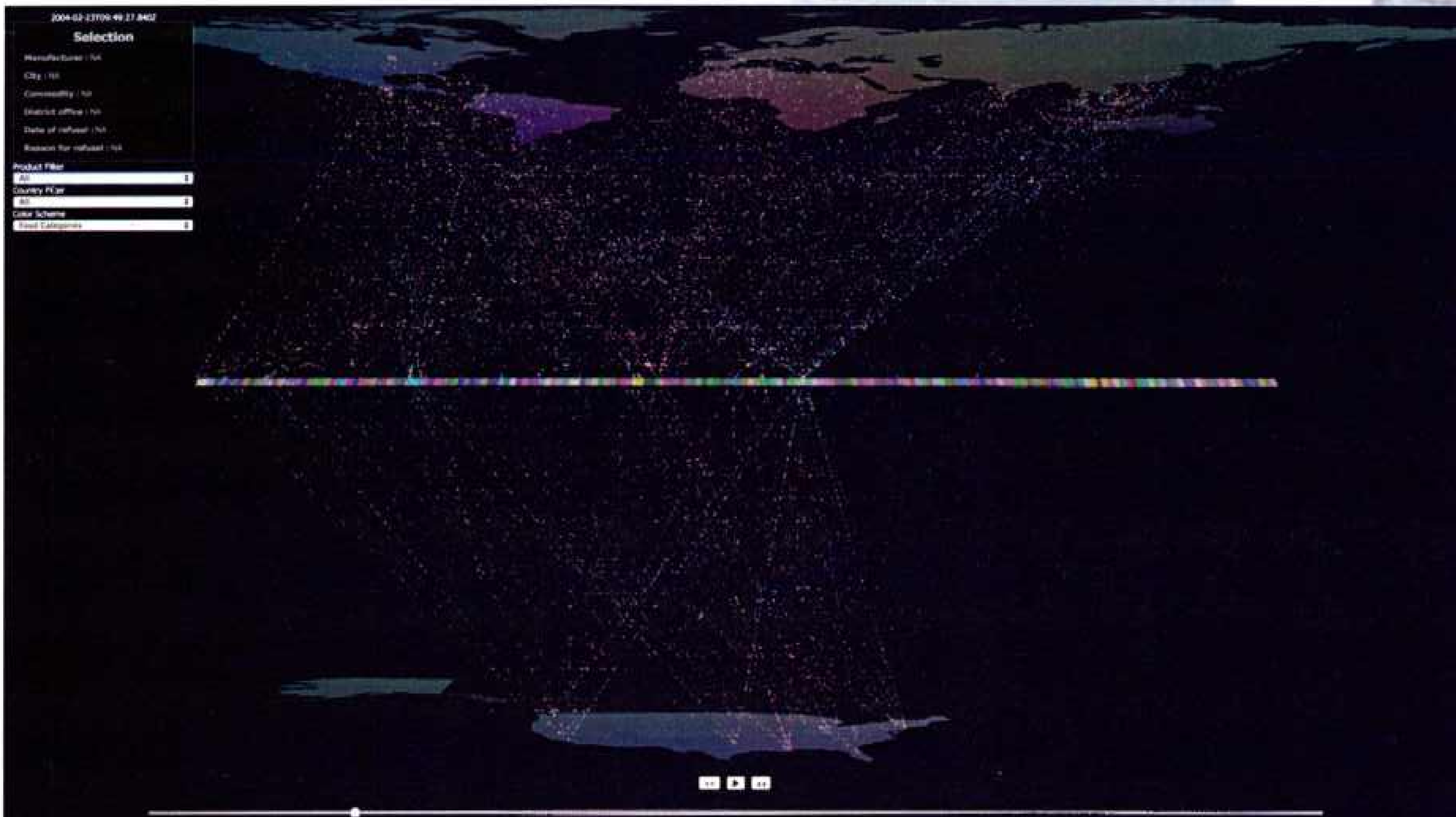


ihr

Food Protection and Defense

- Economically motivated food adulteration and theft
  - Substitution of cheap ingredients
  - Mislabeling and Trans-shipping
  - Port shopping
  - Disappearances of large meat shipments
- Politically motivated food supply chain infiltration
  - Cyber threats (infiltration, DoS potential)
  - Vulnerability assessment and exploitation (scanning, exfiltration, covert attacks on quality assurance mechanisms/processes)

ihmc



ihmc

Summary and Next Steps

- Intelligent visualization can be used to promote sensemaking
- Performance models help focus people on what is important or anomalous
- Agents, policies, learning, and adaptation leverage the power of computation in complex, high tempo environments
- Support for coactive emergence allows humans and agents to quickly converge on threat interpretations
- Support for rapid knowledge sharing through agent libraries, CogLog, and Live Advisories
- With NCFPD, partnering in data sharing and joint tool development

ihmc

Rethinking Decision Making

Dr. Marvin S. Cohen
Perceptronics Solutions
Falls Church, VA 22042

NDM 2015

Questions about NDM

1. Do we know what “decision making” is?

2. What do we have against formal models?

3. Does NDM offer any normative insight other than “This is what experts do”?

How can we make a difference in the real world if:

- We don’t know what class of behaviors we are addressing
- We eschew quantitative predictive / explanatory models
- We have no generally applicable normative framework

1. Do we know what "decision making" is?

Conventional definition: *Selection of an option from a set of alternatives*

But real-world research suggests: Choice is rarely the most important part of a decision process

NDM goes further!

Choice may not occur at all.

Evidence = Self reports that "We don't make decisions."

Naturalistic
DECISION
Making??

- What do deliberative choice *and* rapid recognition have in common, while excluding sneezes and stumbles?

Counterfactual Choice?

- Klein (1998): A decision points occurs when "reasonable options exist, and the commander might have selected a different option. …even if no other option was consciously considered…"
- It's a decision if it *might have been a choice* under different circumstances.
- Problems
  - Sheds no light on the process that actually occurred – *since an alternative played no actual role*.
  - If pattern recognition is a decision, what would *not* count as a decision. Includes too much: habits, innate behaviors, movements induced by electrical stimulation.
- So, does NDM lack a definition of its subject matter?

An overarching paradigm

DM AS COMMITMENT CHANGE

Decision = Change in Commitment

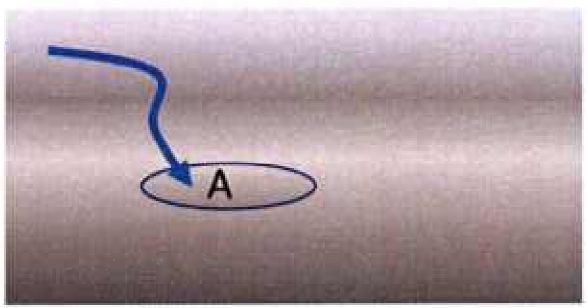
- Evidence for commitment as a psychological state
  - Prospective memory
    - Intention improves memory of previously learned material, advantage disappears after execution (McDaniel et al).
  - Pre vs Post Mindset
    - Vary with respect to attention, memory, biases re success (Gollwitzer)
    - Reasons have different effect before and after consolidation (Baumeister, Montgomery)
  - Inertia
    - Escalation of commitment despite negative initial results (Staw)
- Consistent with *function* of decisions:
  - Provides stability and predictability necessary for preparation, coordination, attention to other tasks
  - Explains resistance to change

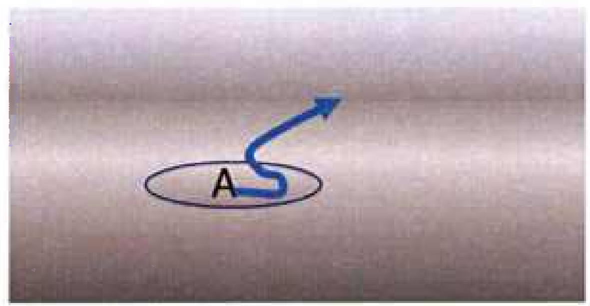
Definition refers to *future* impact rather than details of generating process.

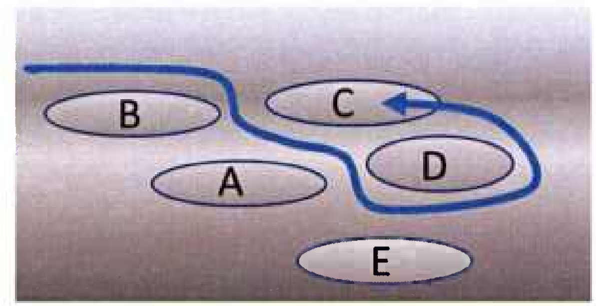
Once Commitment Change is Added, Choice Can Drop Out

A decision *always* has alternative possible outcomes. But they need not be real alternatives for commitment.

Real choice = More than one actual commitment prospect influences the DM process







"Go / no-go" decision

MATCHING

Change of mind

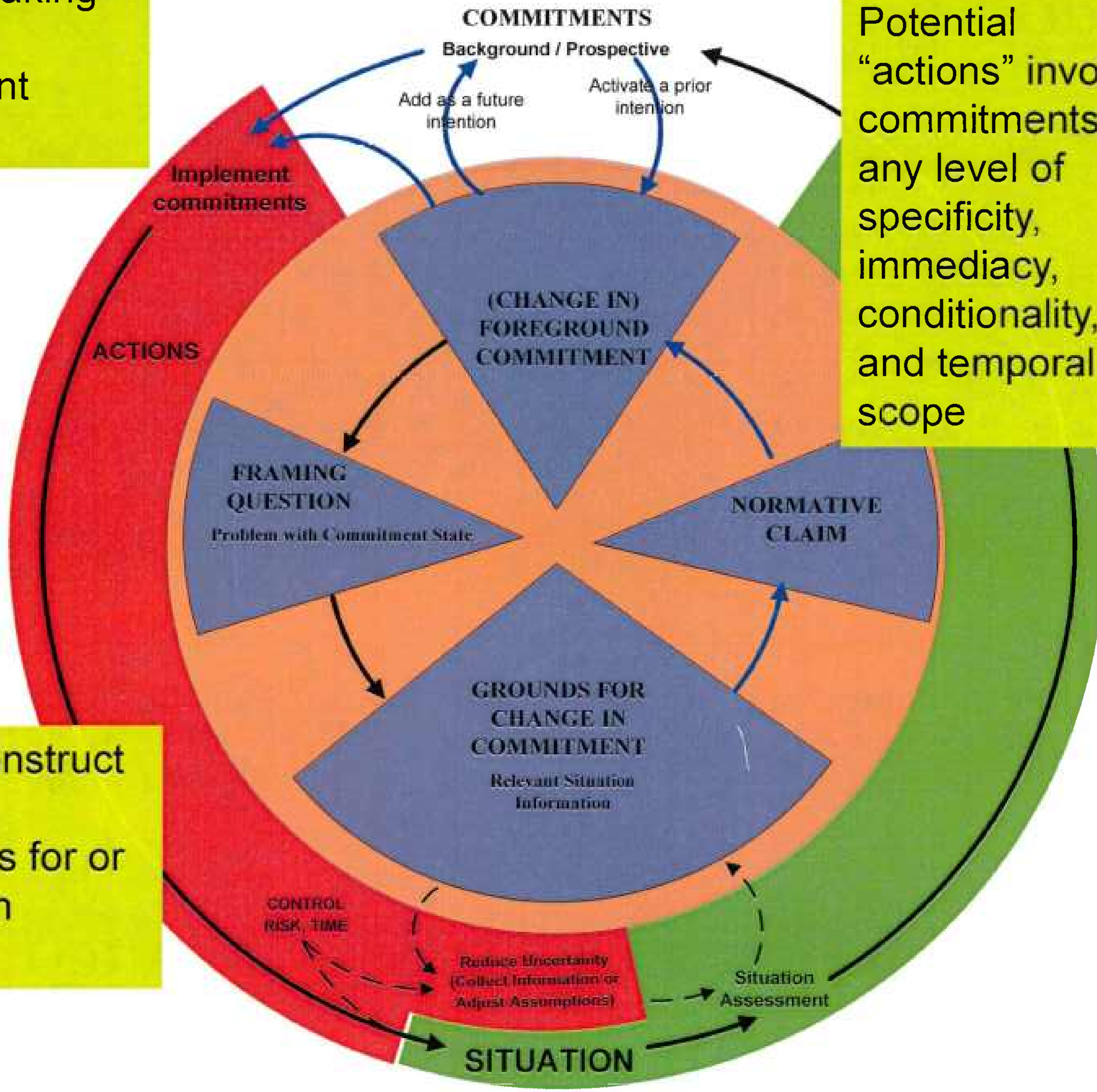
REASSESSMENT

Select among real alternatives

CHOICE

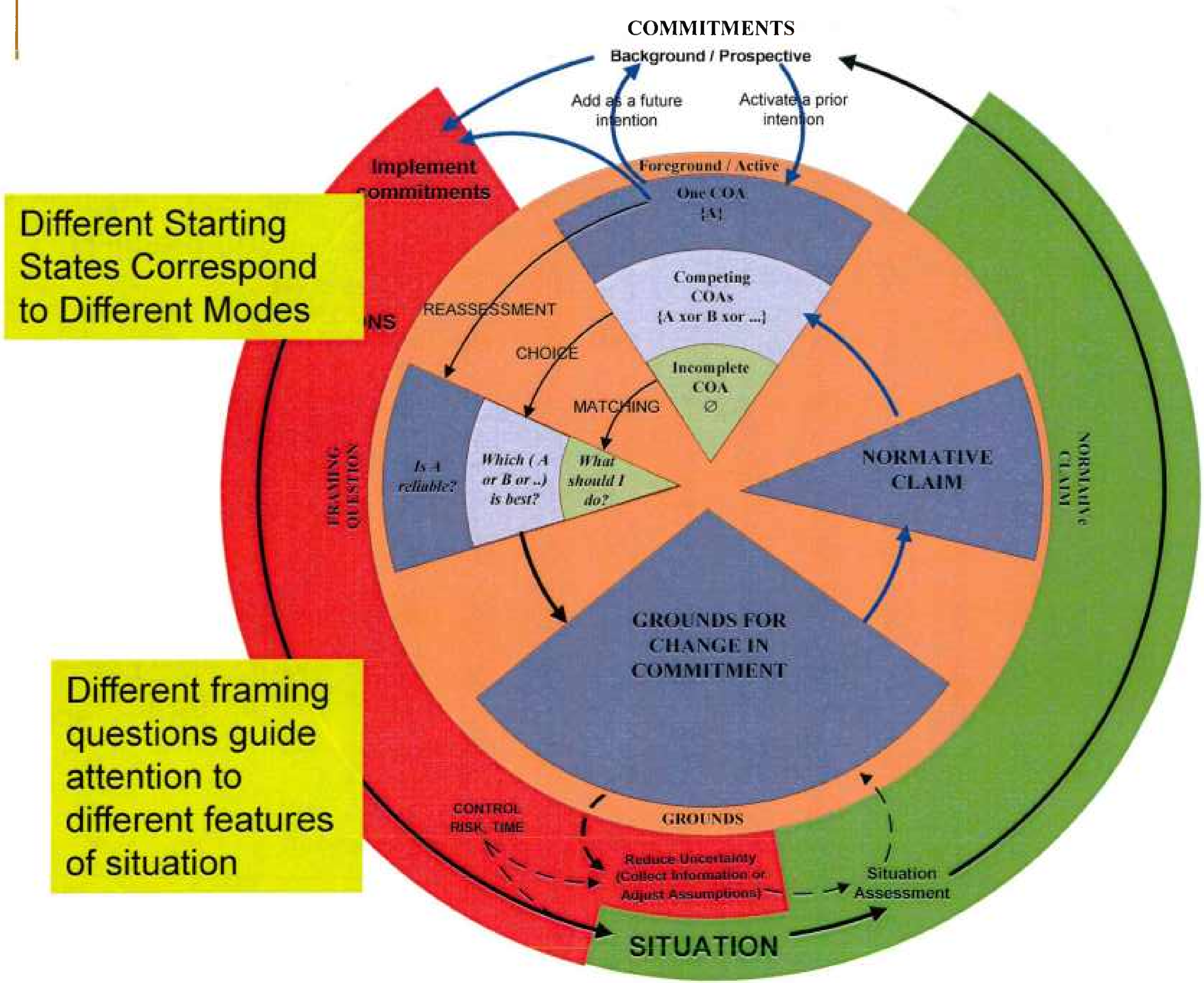
Modes of Commitment Change

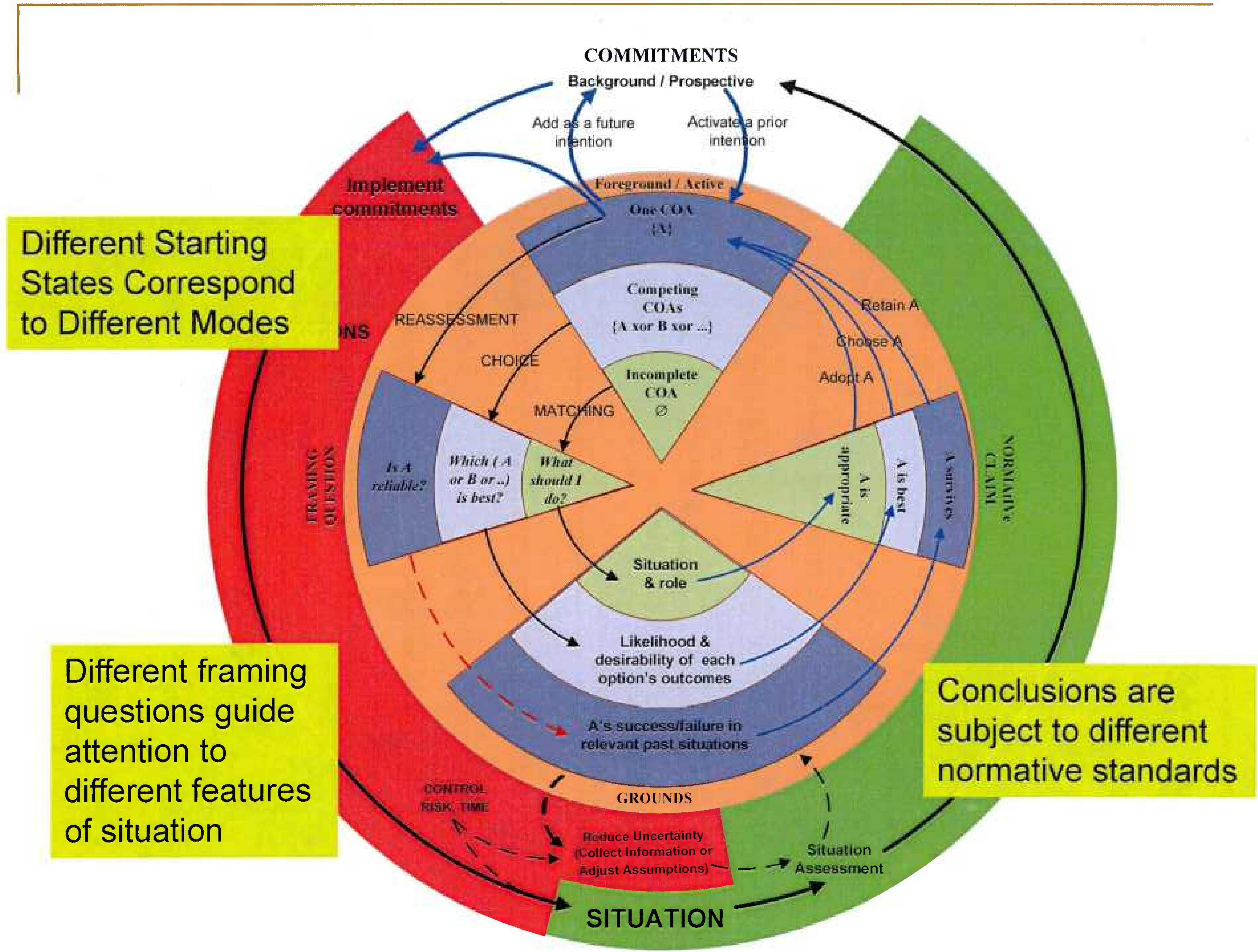
Decision Making = Cycles of Commitment Change



Potential "actions" involve commitments at any level of specificity, immediacy, conditionality, and temporal scope

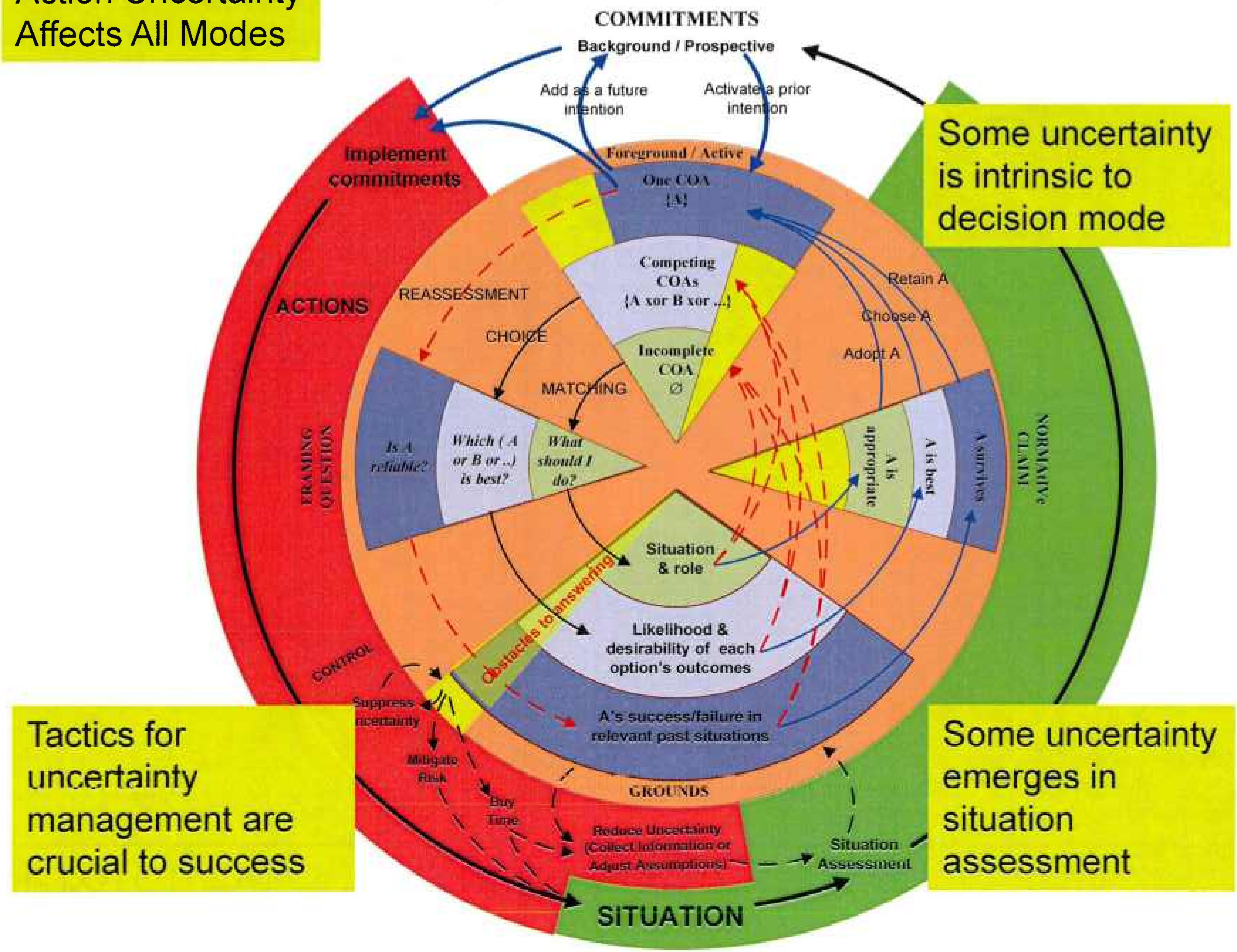
Cycles construct implicit arguments for or against an "action"


COMMITMENTS
Background / Prospective
Add as a future intention
Activate a prior intention
Implement commitments
Foreground / Active
One COA {A}
Competing COAs {A xor B xor ...}
Incomplete COA ∅
REASSESSMENT
CHOICE
MATCHING
Is A reliable?
Which (A or B or ..) is best?
What should I do?
FRAMING QUESTION
NORMATIVE CLAIM
NORMAtIVe CLAIM
GROUNDS FOR CHANGE IN COMMITMENT
GROUNDS
CONTROL RISK, TIME
Reduce Uncertainty (Collect Information or Adjust Assumptions)
Situation Assessment
SITUATION
Different Starting States Correspond to Different Modes
Different framing questions guide attention to different features of situation

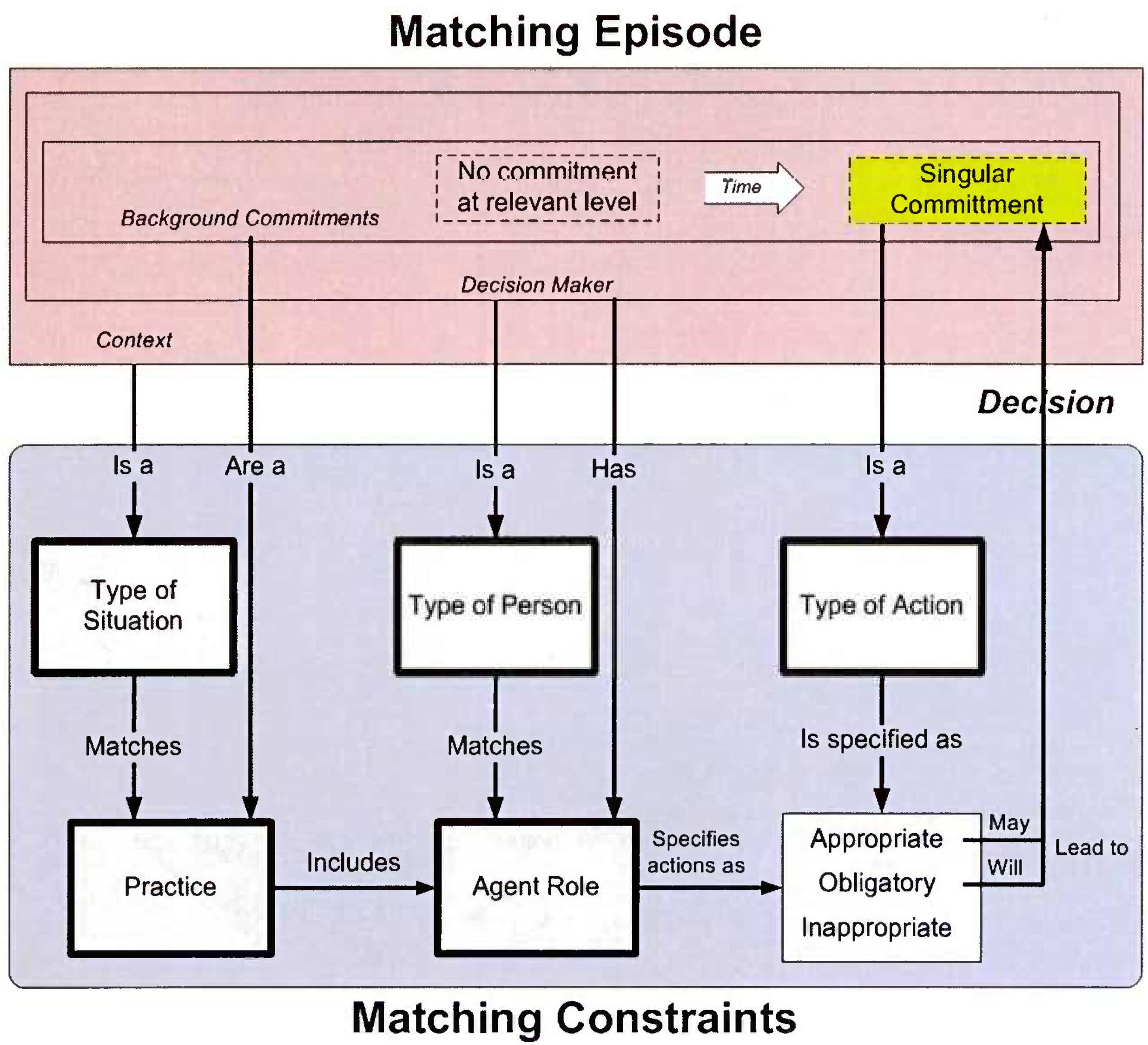


COMMITMENTS
Background / Prospective
Add as a future intention
Activate a prior intention
Implement commitments
Foreground / Active
One COA {A}
Competing COAs {A xor B xor ...}
Incomplete COA ∅
Different Starting States Correspond to Different Modes
REASSESSMENT
CHOICE
MATCHING
Retain A
Choose A
Adopt A
FRAMING QUESTION
Is A reliable?
Which (A or B or ..) is best?
What should I do?
A is appropriate
A is best
A survives
NORMATIVE CLAIM
Situation & role
Likelihood & desirability of each option's outcomes
A's success/failure in relevant past situations
Different framing questions guide attention to different features of situation
Conclusions are subject to different normative standards
CONTROL RISK, TIME
GROUNDS
Reduce Uncertainty (Collect Information or Adjust Assumptions)
Situation Assessment
SITUATION

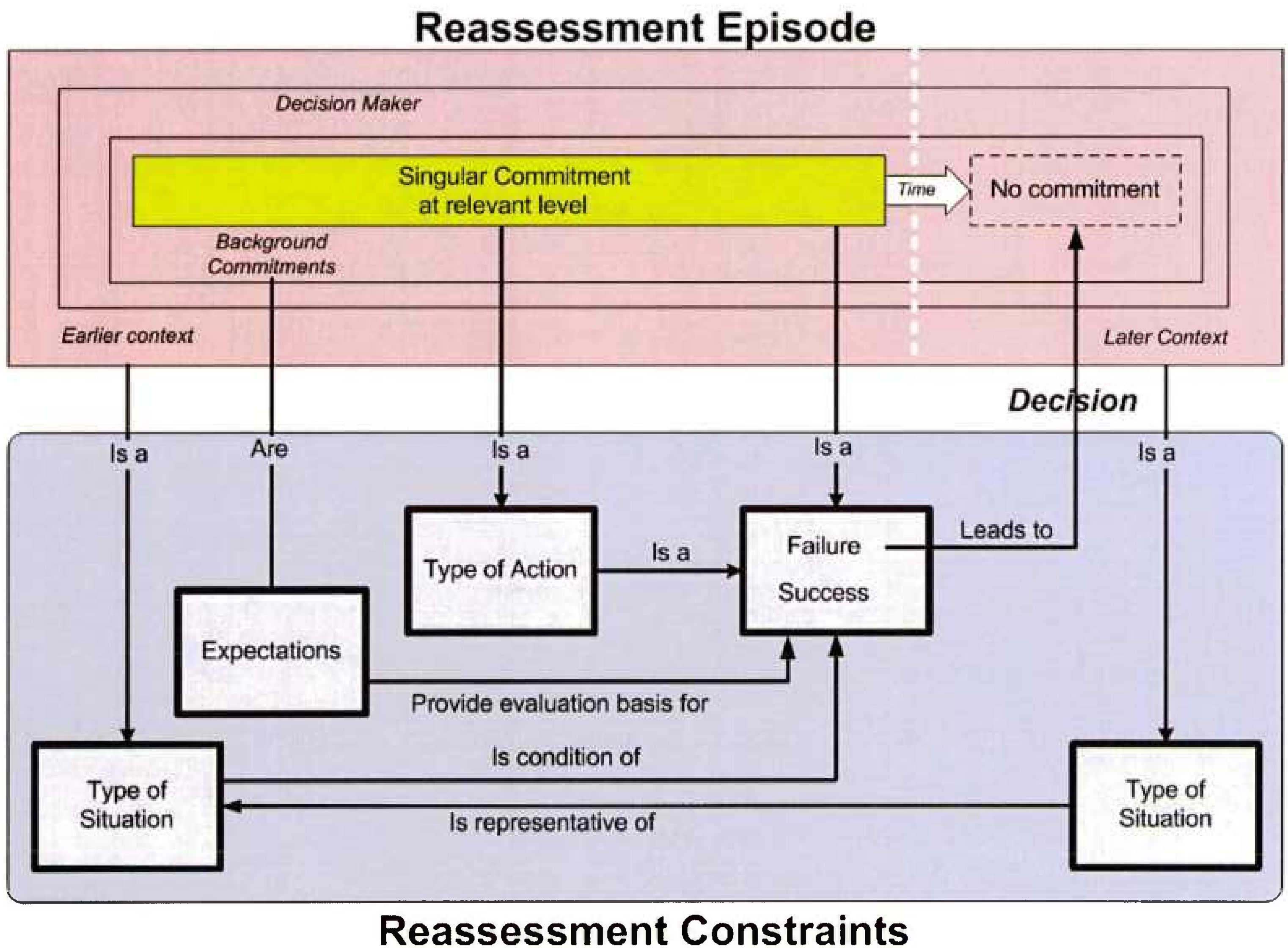
Action Uncertainty Affects All Modes



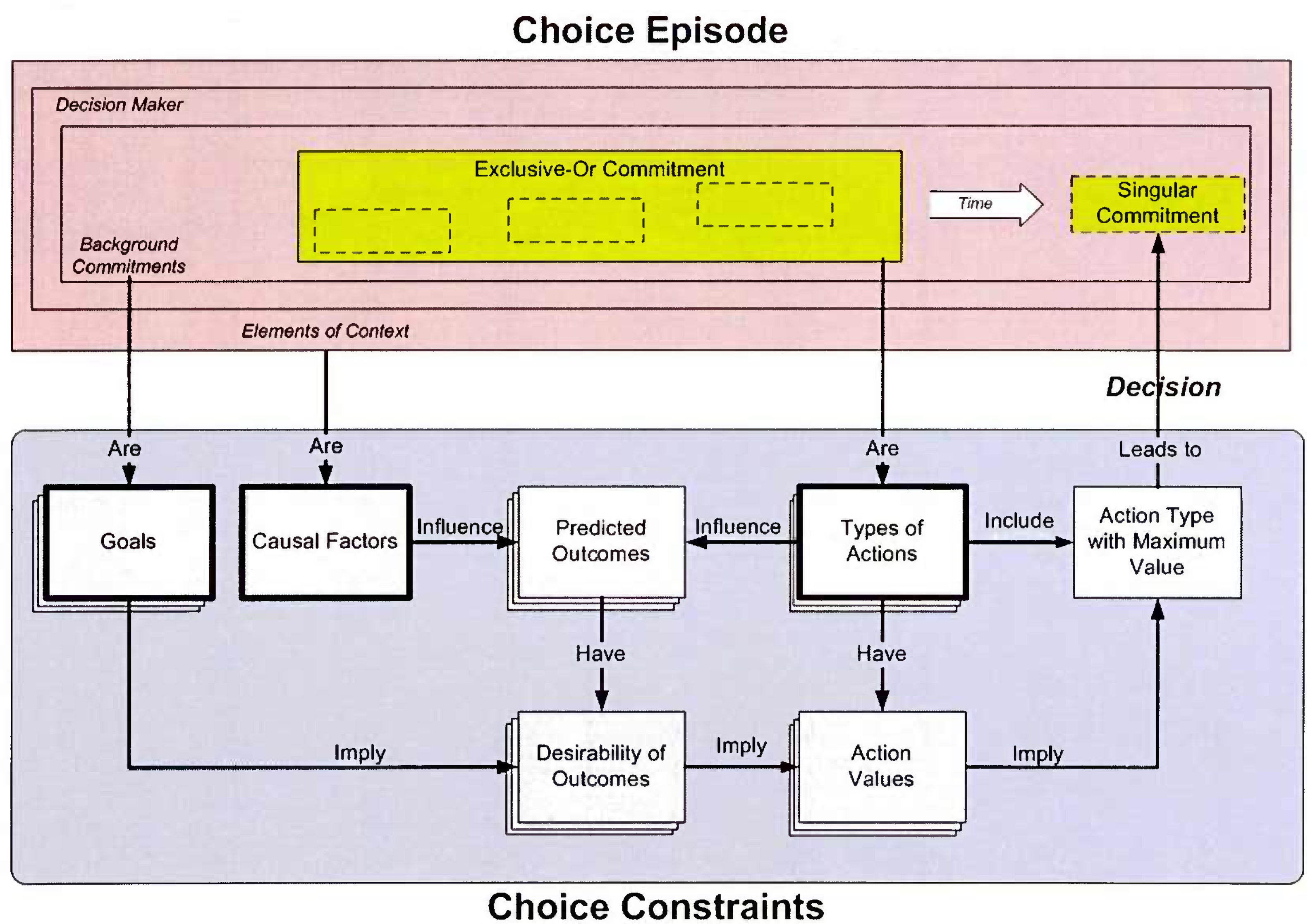
Prototype Schema #1



Prototype Schema #2



Prototype Schema #3



Partial Answer to (1): *What is A Decision?*

- DM is any process that can cause changes in action commitment.
  - Includes adding, selecting, or rejecting commitment possibilities.
  - DM can & does occur with only one live commitment possibility.
  - Experienced as a decision to the extent that uncertainties must be overcome.
  - Different normative principles apply to different modes of commitment change.

2. What do we have against formal models?

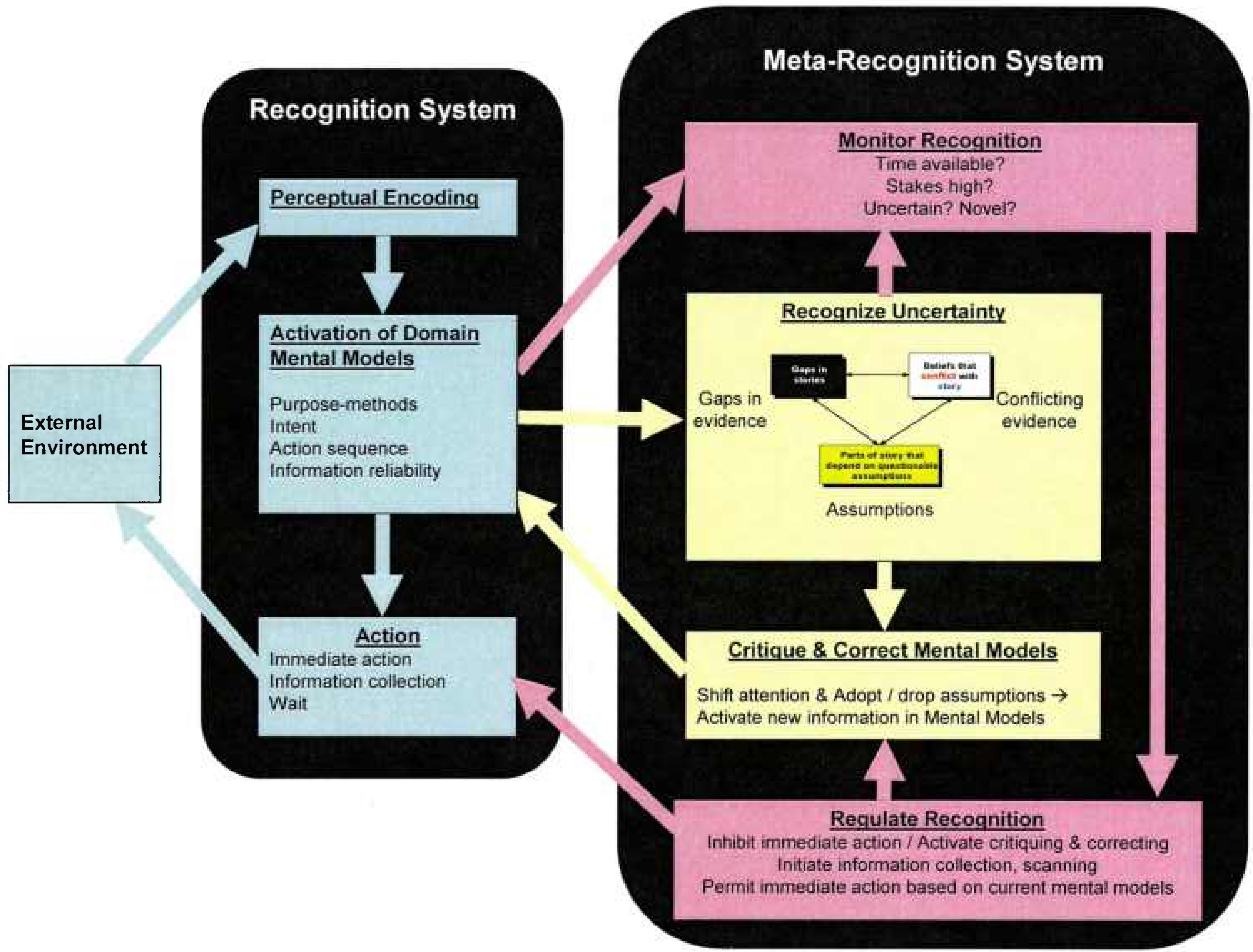
- Can / should we back up our informal models and "macrocognitive" concepts – such as pattern recognition, situation awareness, story-building, & decision making – with a more complete account?
    - What are "units of knowledge" – such as schemas, frames, rules, and mental models? Are they crisply bounded and well-defined in advance of use?
    - Are differences in proficiency entirely due to knowledge? Does the *how* as well as the *what* contribute?
    - What about goals, preferences, motivation, social context, self-regulation?

RECOGNITION / METACOGNITION

An informal mantra:

First recognize. Then, if time & stakes permit, verify, elaborate, modify!

Critical Thinking about Recognition



Uncertainty in stories

- When time, stakes, and uncertainty afford deliberation,
  - DM’s create a story around a key assessment or plan
  - Think critically about the story
- Different types of uncertainty are addressed by different tactics.
  - Collecting information to fill gaps
  - Adjusting assumptions to fill gaps and resolve conflict
  - Evaluating plausibility of the result to manage assumptions
- Correcting a problem alters the story.
- The process is iterative: Corrections can (but need not) lead to other problems with the story, causing it to further evolve!

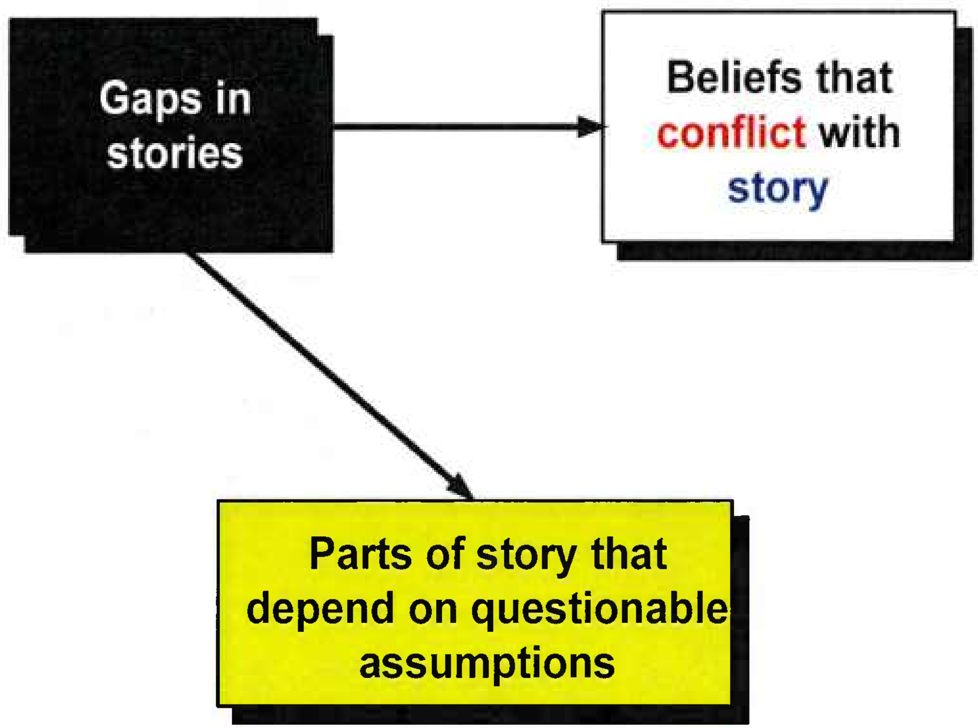
Uncertainty in stories

Filling gaps can lead to conflicting information or assumptions

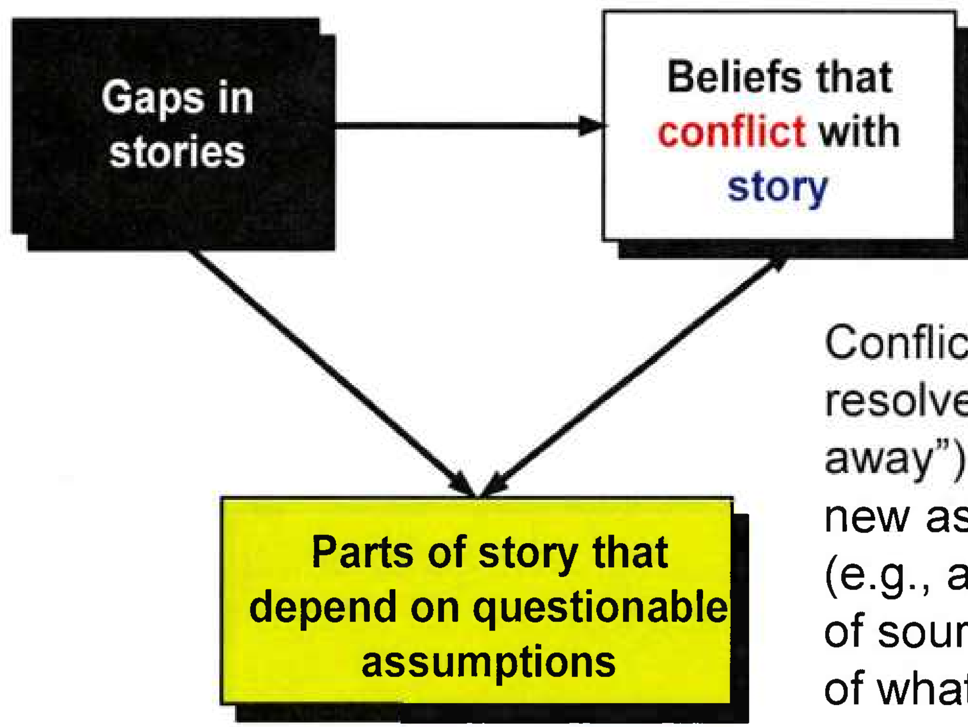
Gaps in stories

Uncertainty in stories

Filling gaps can lead to conflicting information or assumptions

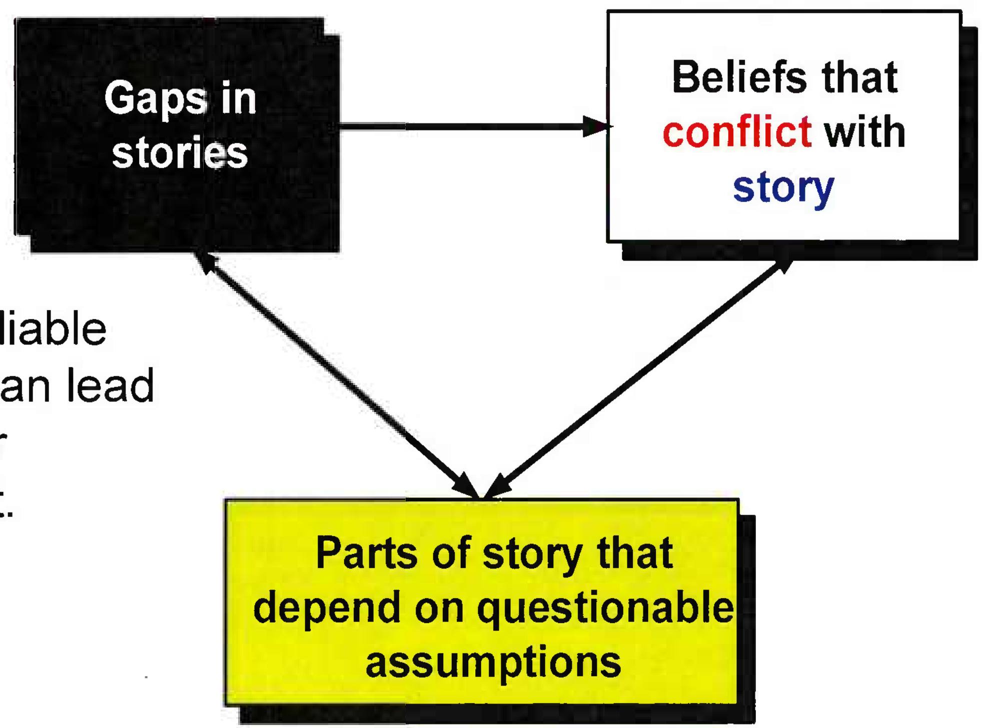


Uncertainty in stories



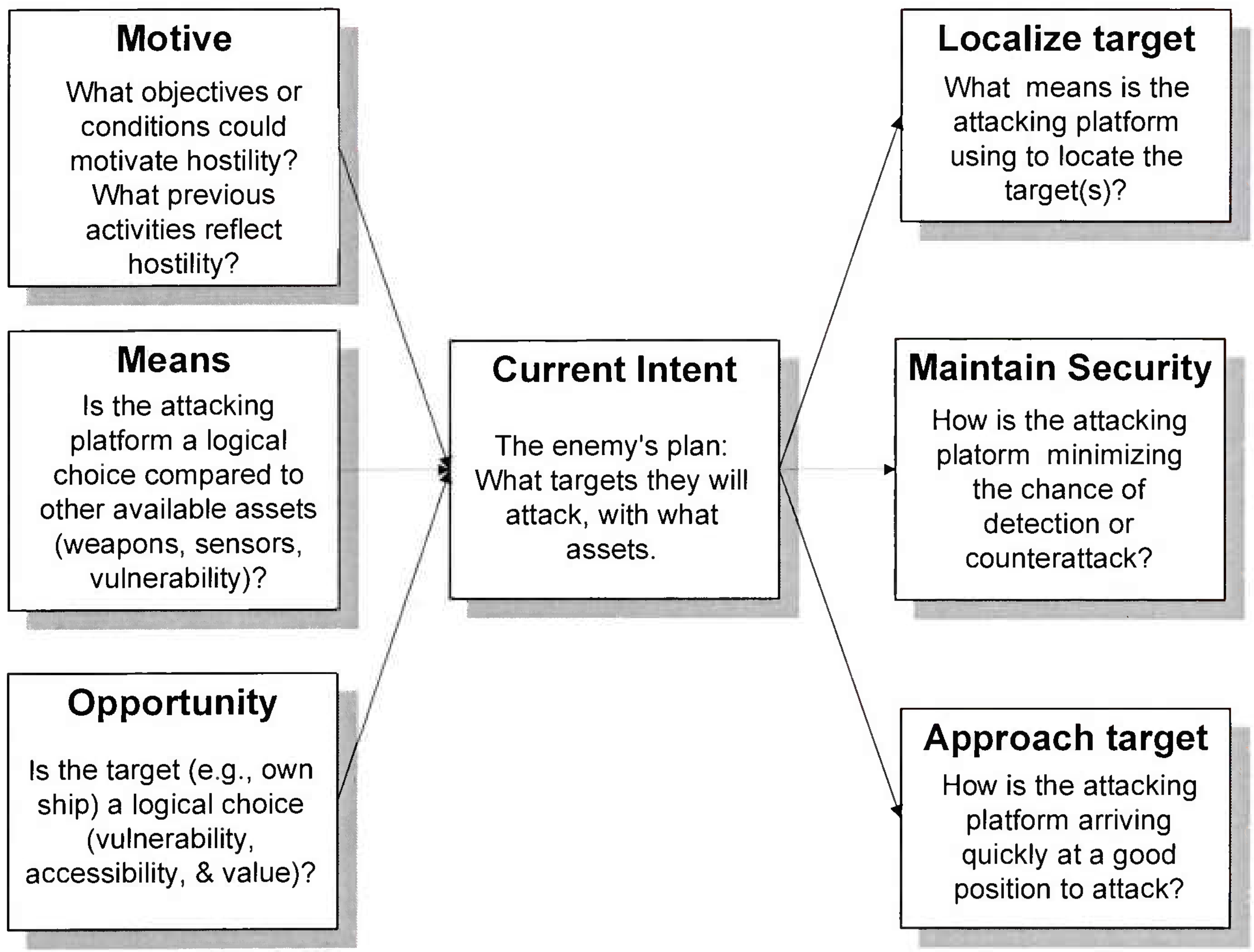
Conflict can be resolved ("explained away") by adopting new assumptions (e.g., about credibility of sources or meaning of what they report).

Uncertainty in stories



Dropping unreliable assumptions can lead to new gaps or restore conflict.

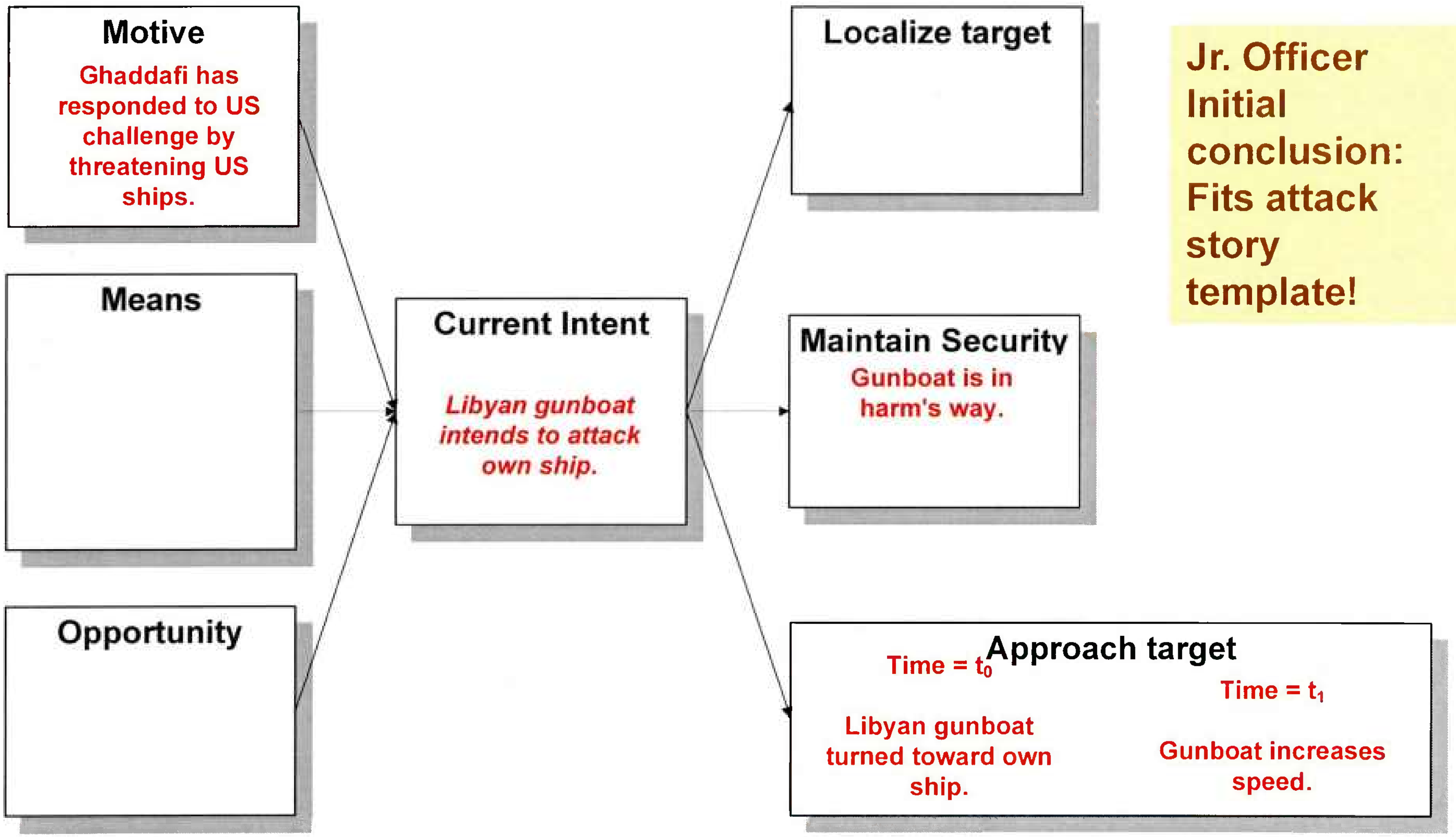
Example: *Intent to Attack Story Schema*



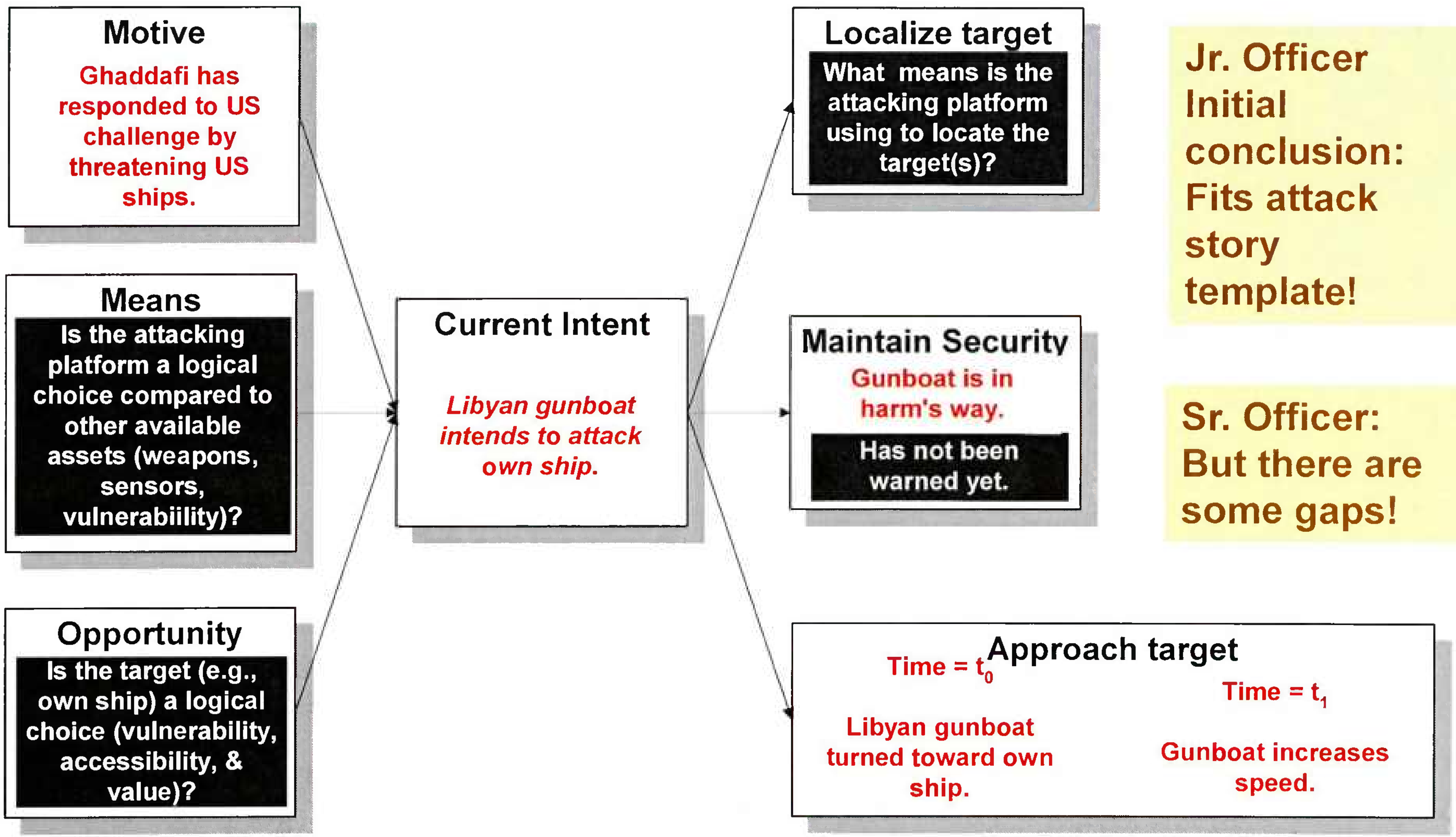
With experience, decision makers learn typical events and causal associations.

They use these expectations to interpret observations – and to test hypotheses!

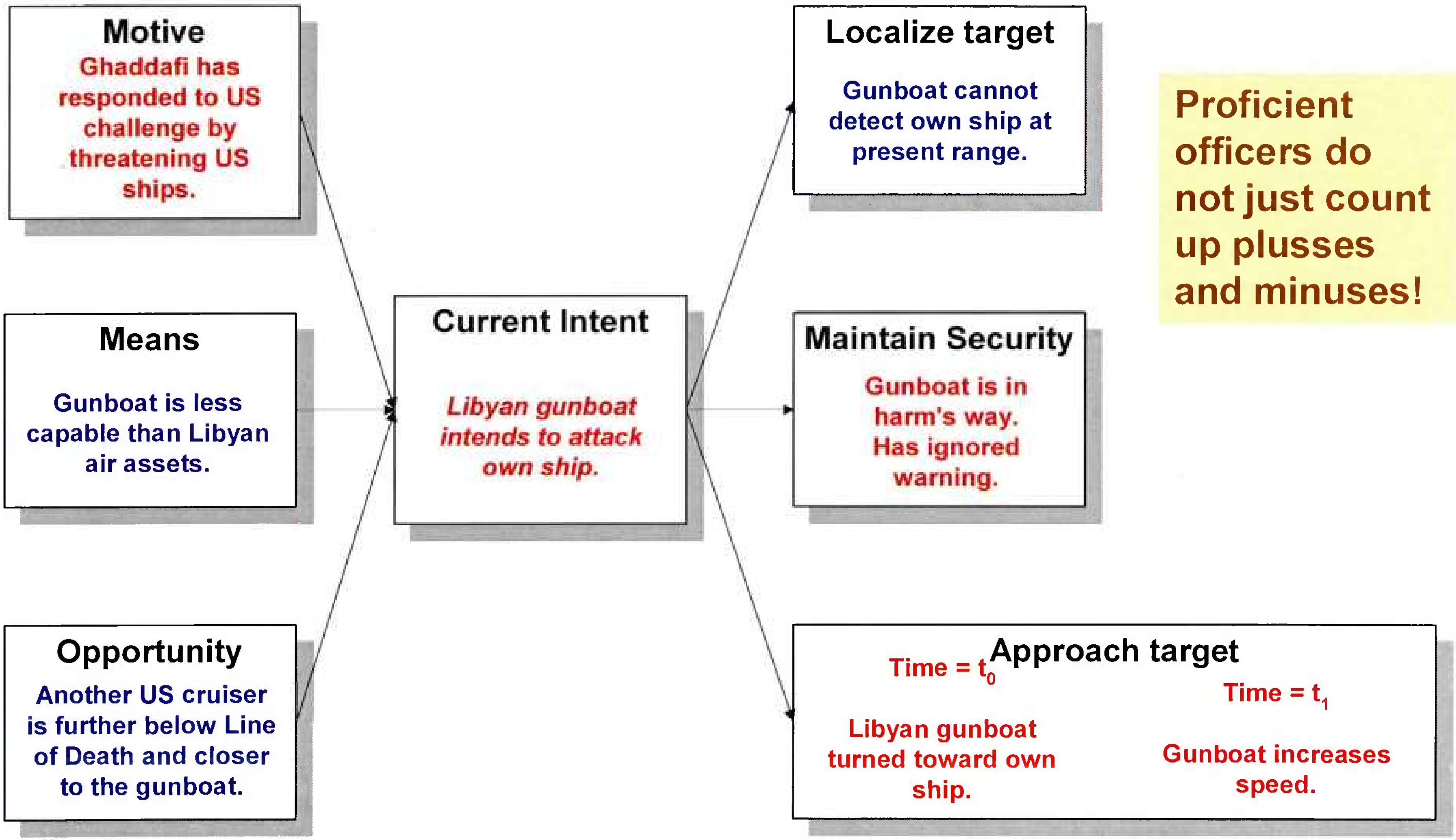
Aegis cruiser spots Libyan gunboat leave port, speed up, and head toward own ship in Gulf of Sidra



Aegis cruiser spots Libyan gunboat leave port, speed up, and head toward own ship in Gulf of Sidra

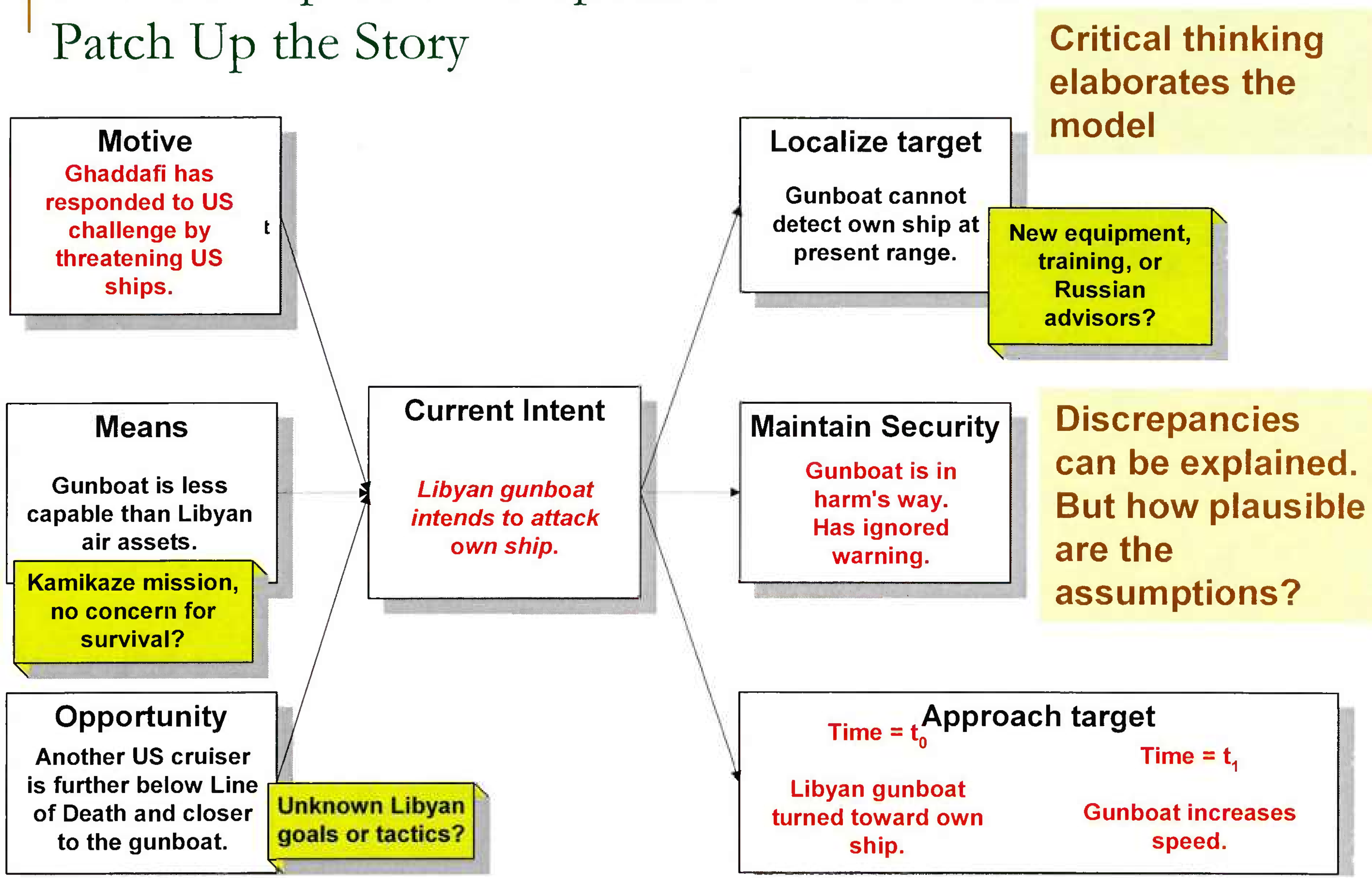


When Gaps are Filled, Get Conflicting Conclusions

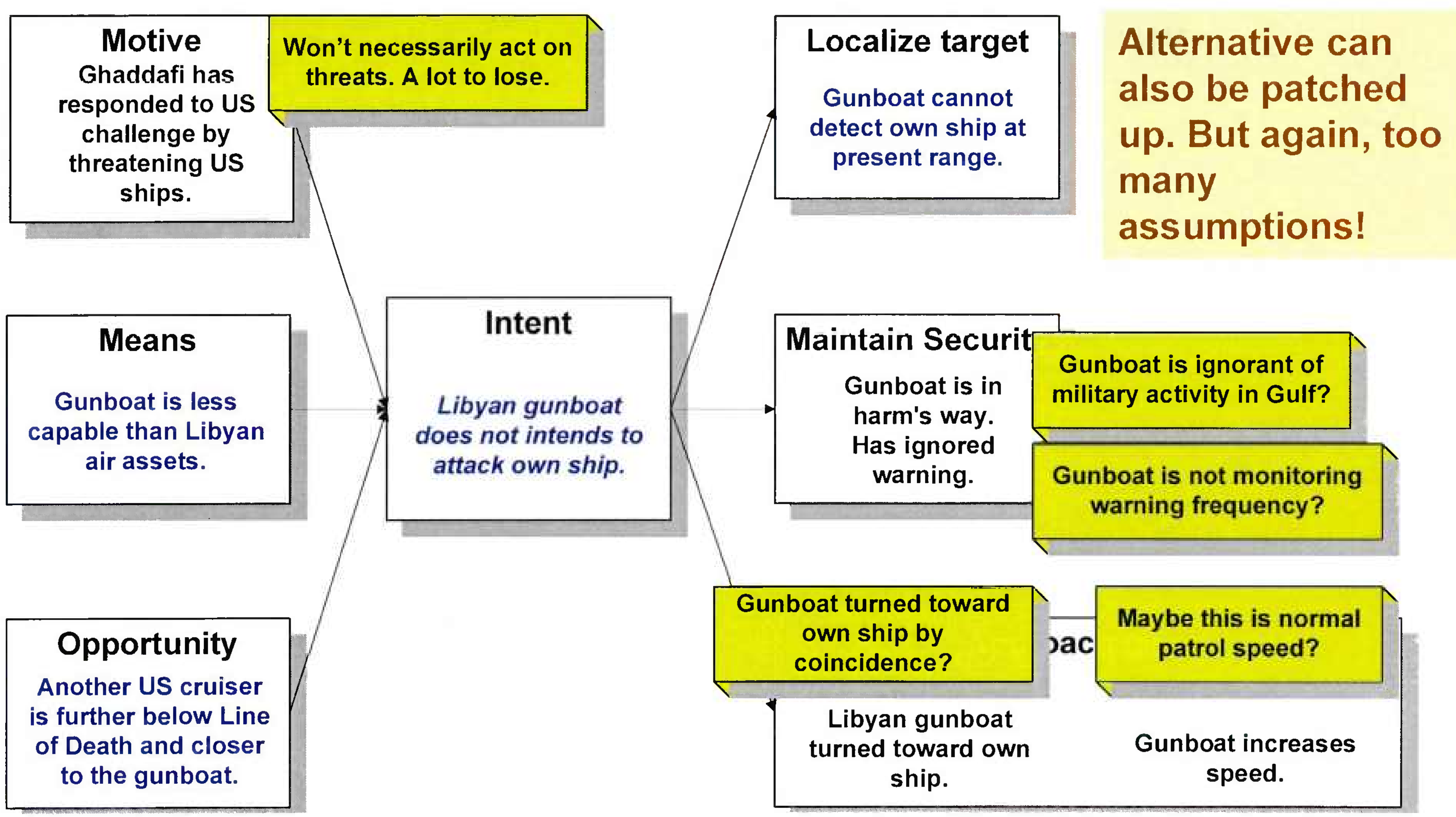


Proficient officers do not just count up plusses and minuses!

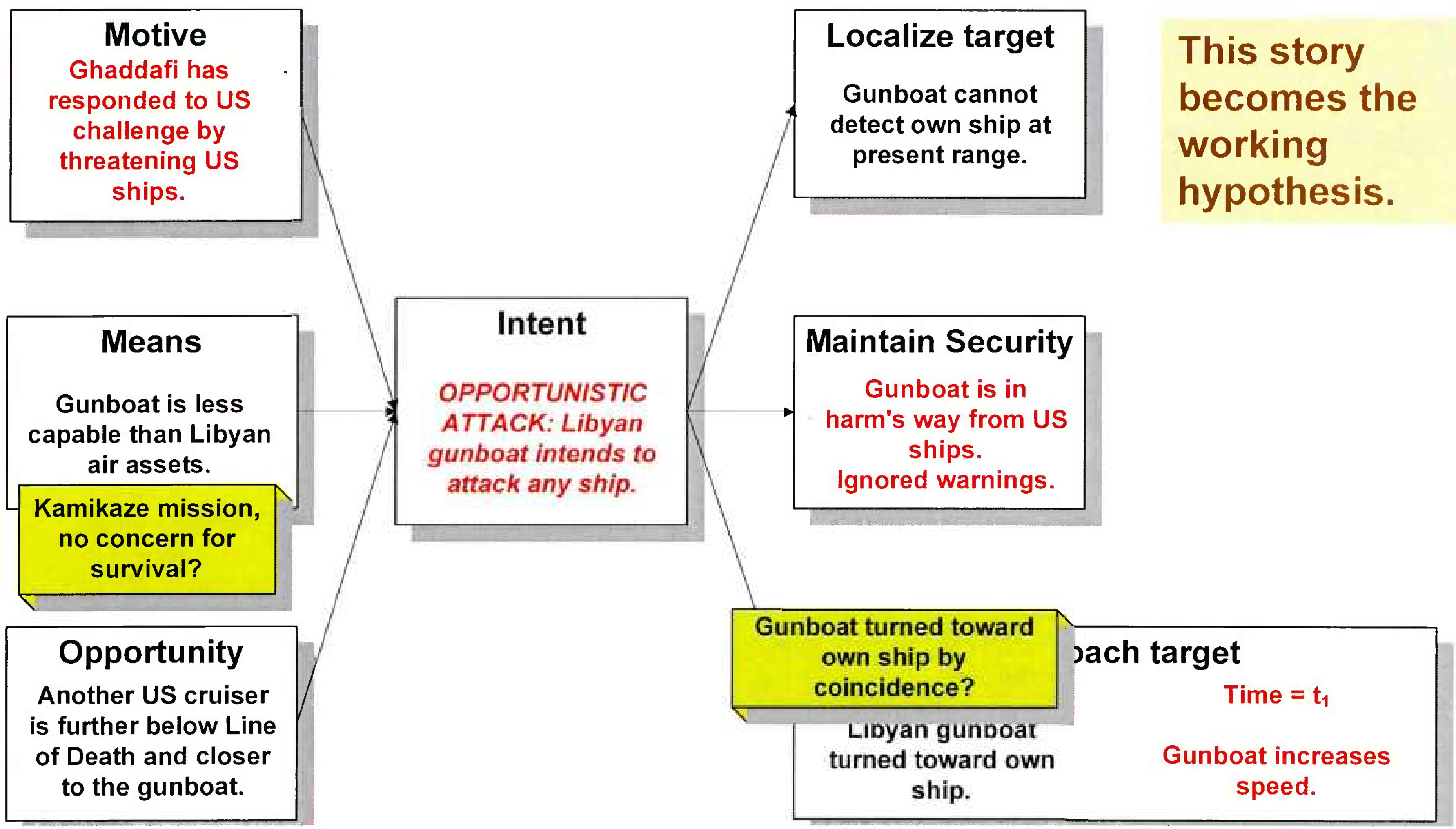
Use Assumptions to Explain the Conflict and Patch Up the Story



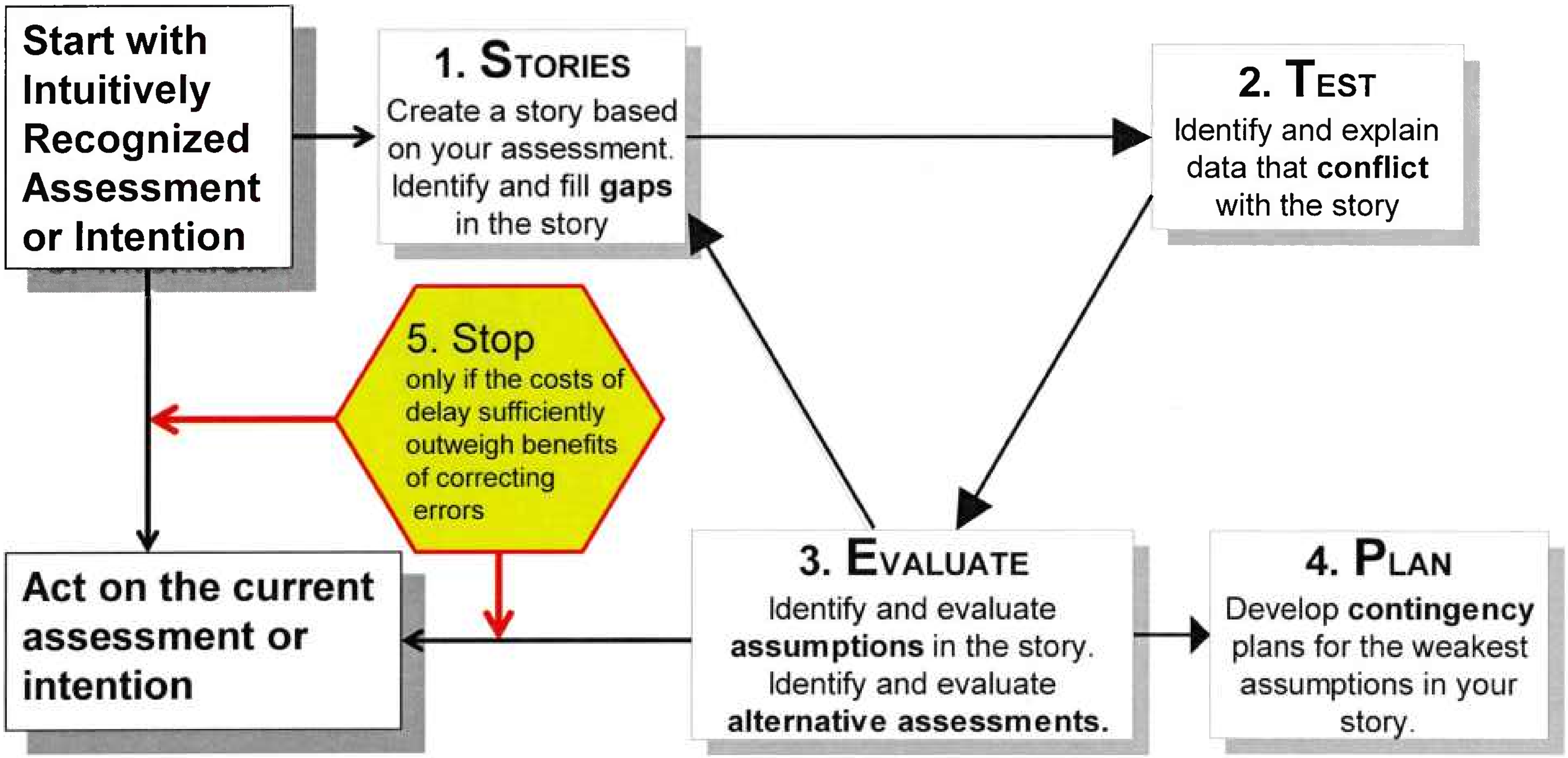
Create Alternative Story, which also has Conflicts



Find an even better story, with fewer assumptions than either of the other two



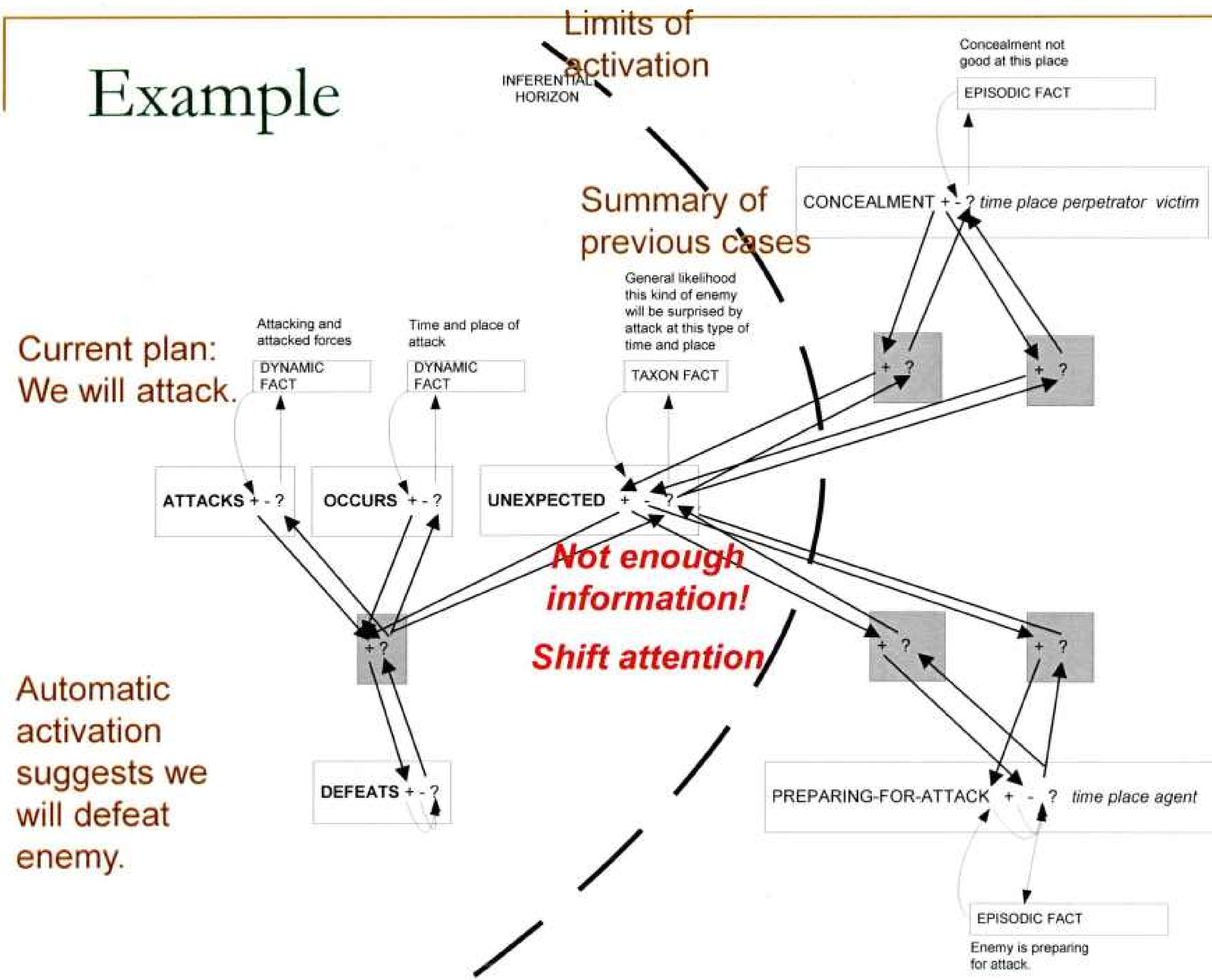
Critical Thinking Training: STEPS



Mix of automaticity and attention

NEURAL MODEL OF KNOWLEDGE INSTANTIATION & ELABORATION

Example

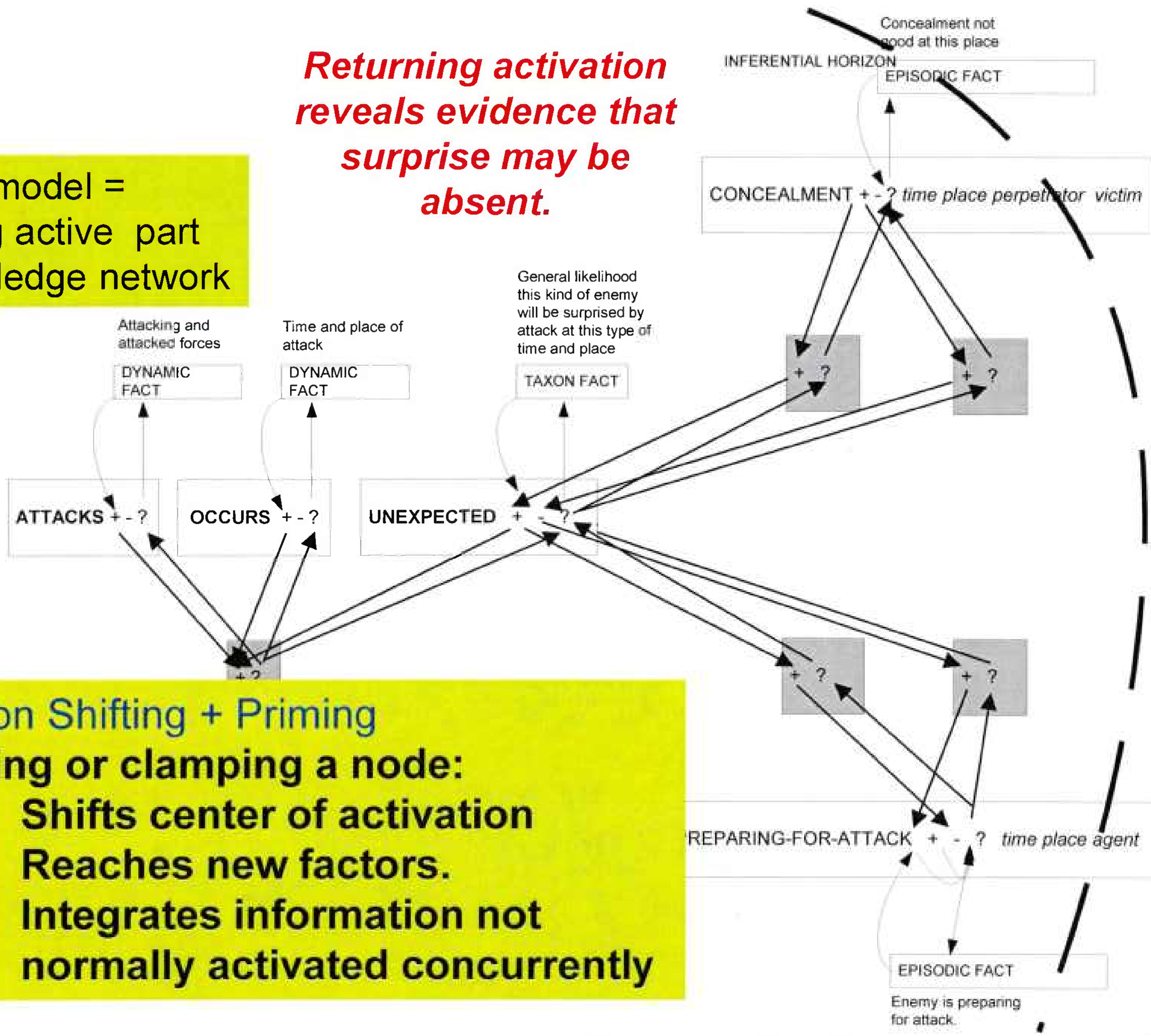


Limits of activation
INFERENTIAL HORIZON
Summary of previous cases
Current plan: We will attack.
Automatic activation suggests we will defeat enemy.
Attacking and attacked forces
DYNAMIC FACT
Time and place of attack
DYNAMIC FACT
General likelihood this kind of enemy will be surprised by attack at this type of time and place
TAXON FACT
Concealment not good at this place
EPISODIC FACT
CONCEALMENT + - ? time place perpetrator victim
ATTACKS + - ?
OCCURS + - ?
UNEXPECTED + - ?
DEFEATS + - ?
Not enough information!
Shift attention
PREPARING-FOR-ATTACK + - ? time place agent
EPISODIC FACT
Enemy is preparing for attack.

CTI

Returning activation reveals evidence that surprise may be absent.

Mental model = evolving active part of knowledge network



Attention Shifting + Priming

Querying or clamping a node:

1. **Shifts center of activation**
2. **Reaches new factors.**
3. **Integrates information not normally activated concurrently**

Neural R/M Model

- Commitment involves a self-reinforcing cycle of activation
  - Stability requires return activation
  - Loops are closed by perceptions, memories, actions, values
  - Prompts immediate action or generates intention
- Critiquing and Correcting
  - Presupposes commitment
  - If successful, widens sphere of self-reinforcing activation → increases robustness of commitment

Some Subtleties

QUICK TEST FOR TIME MANAGEMENT

Quick Test for Delaying Irreversible Commitment

Simple Decision Model

Delay irreversible commitment only if:

$$\frac{\text{CostOfDelay}}{\text{CostOfError}} < \text{Commitment} * \Delta\,\text{ProbOfError}$$

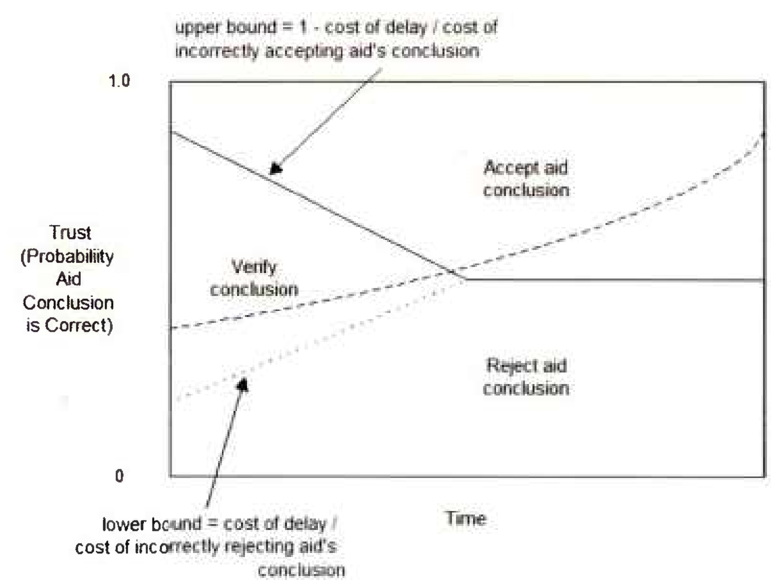
$$= \text{Commitment} * (\text{UsualTrustInThisDomain} - \text{TrustInConclusion})$$

Potential improvement from vetting action

Motivation to vett

Plan or hypothesis is current default

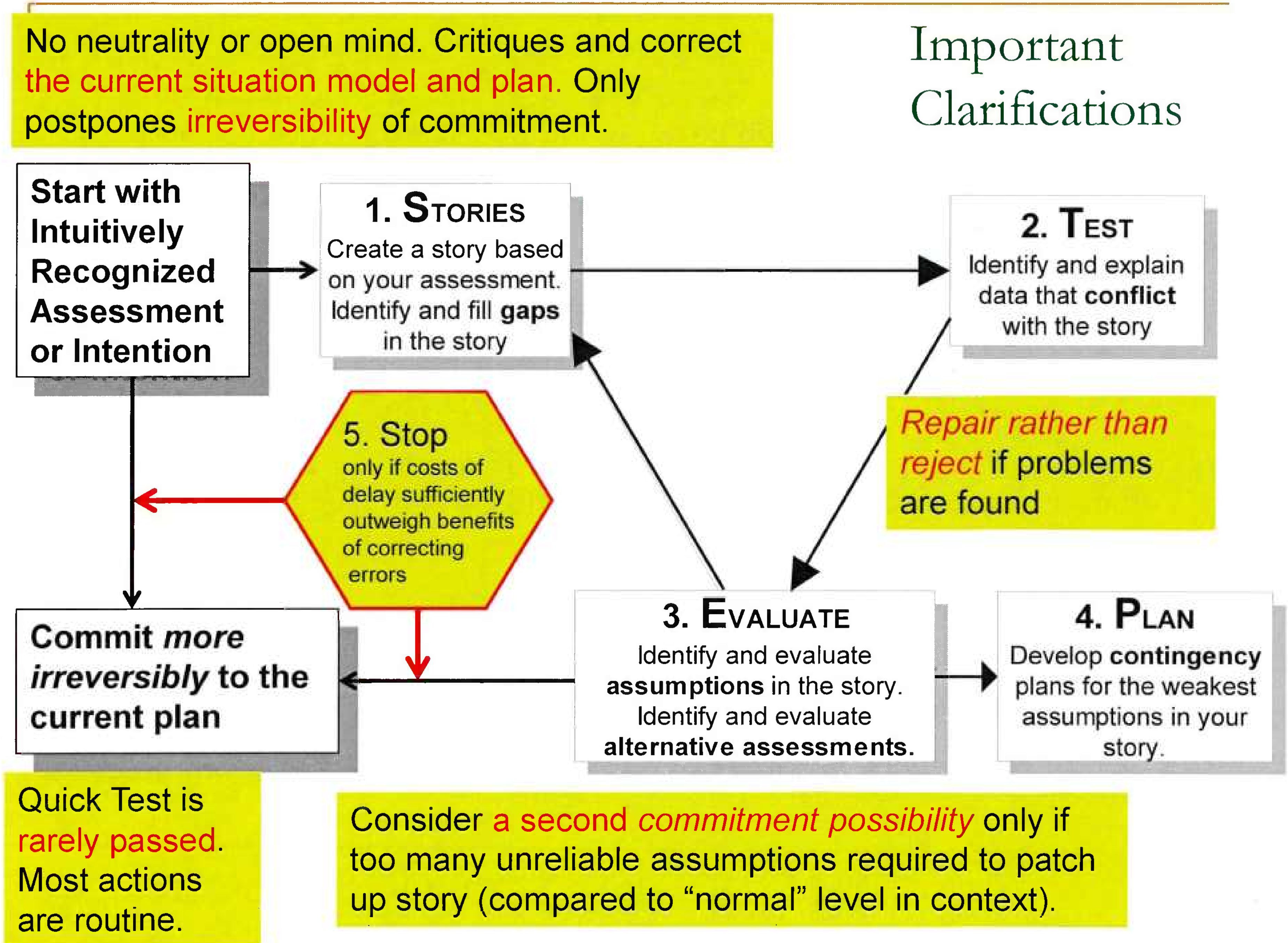
Trust is compared to a "normal" level



Important Misunderstandings: Klein's (2007) Criticism of R/M

- "…efforts to train decision makers to keep an open mind (e.g., Cohen [et.al.]) can be counterproductive."
- "We hypothesize that methods designed to prevent premature commitment to a frame (e.g., the recognition /metacognition approach…) will degrade performance under conditions where active attention management is needed (using frames) and where people have difficulty finding useful frames."

Important Clarifications



Klein's Proposed Stopping Rule

Klein et al.: "sensemaking usually ceases when the data and frame are brought into congruence..."

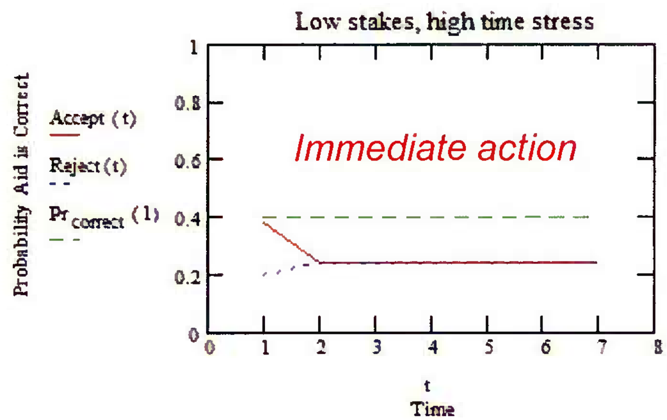
Does not work as a "stopping rule"!

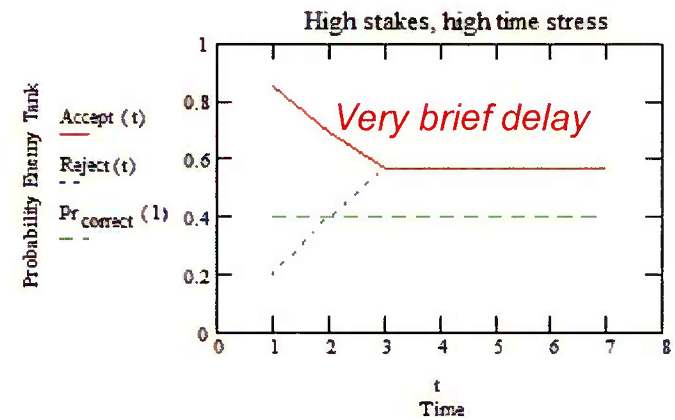
- Not necessary: Deliberation may stop despite *lack of congruence*
  - People (rightly!) may act on well-supported hypotheses despite some gaps and conflicts, even if early efforts fail
  - *Must consider **Cost of delay*** and ***Normal level of uncertainty***!
- Not sufficient: Decision making may continue despite *congruence* (e.g., in novel situations).
  - DMs actively probe for uncertainty when time is available and stakes are high. There is always the possibility of discovering relevant new information
  - Must consider ***Stakes*** and **open-ended possibility of relevant new information**!

Quick Test Makes Predictions

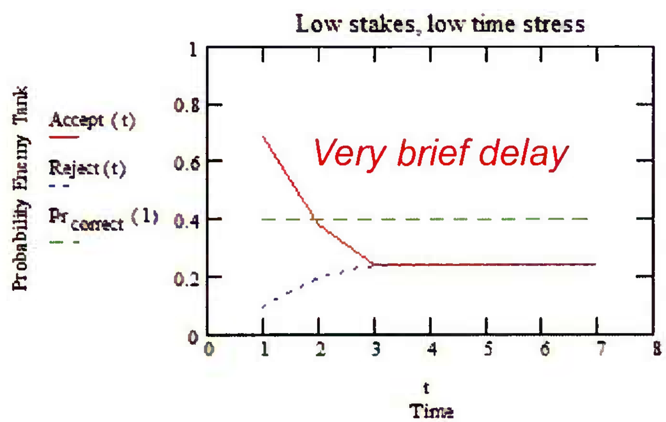
Low Stakes | High Stakes

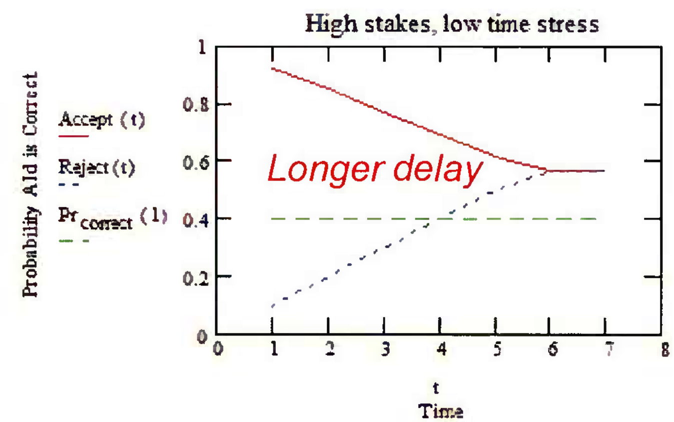
High Time Stress



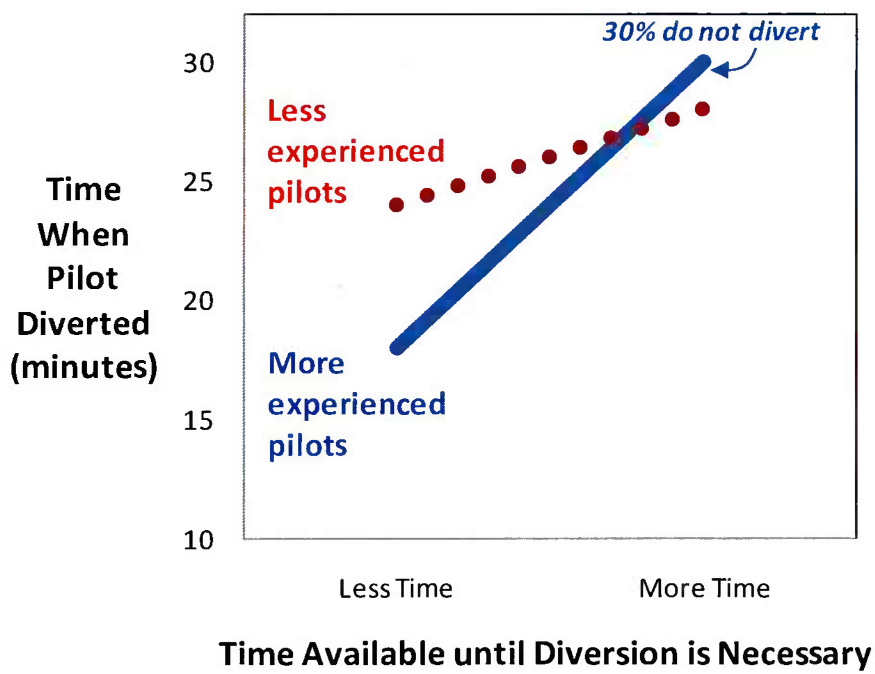


Low Time Stress

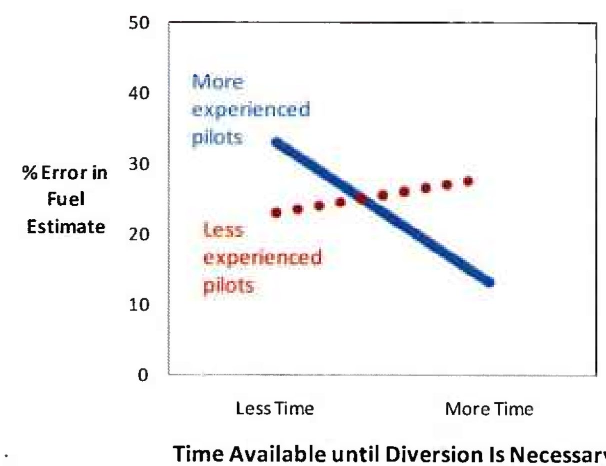




Use of Time for DM is an Experience-Based Skill



Experienced pilots regulated the amount of time taken for decision making.
Less experienced pilots did not.



Differential awareness of available fuel was not responsible.

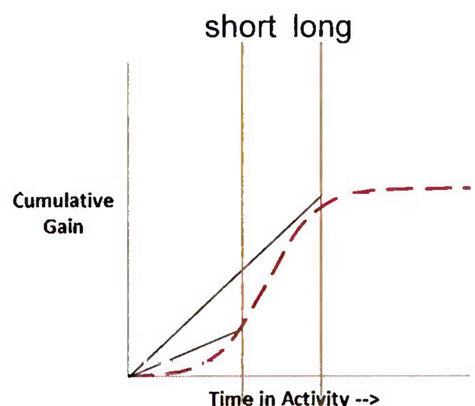
Sampling Can Cause Bias

Delay irreversible commitment only if: $\frac{\text{CostOfDelay}}{\text{CostOfError}} < \Delta\,\text{ProbOfError} \approx 1 - \text{TrustInConclusion}$

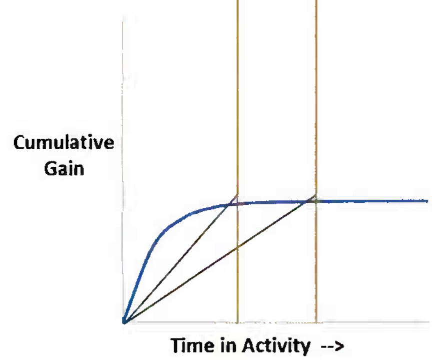
*Quick Test inputs depend on **previously sampled intervals***

Suppose:

Deliberation requires long interval sample to return benefit.



Advantage of quick action vanishes after short interval



RESULT:
Exposure to limited sample of durations can trap decision makers in less productive activity.

Cumulative gain / duration = efficiency

What the Formal Model Highlights

- Neural Implementation
  - Knowledge representation is open-ended, dynamic
  - Attention management is a skill
- Quick Test for Deliberation
  - Self-regulation is necessary because of open-ended character of critical thinking
  - Uncertainty must exceed what is “normal” in the context to motivate deliberation
  - Motivation (cost of delay, cost of errors) must be involved

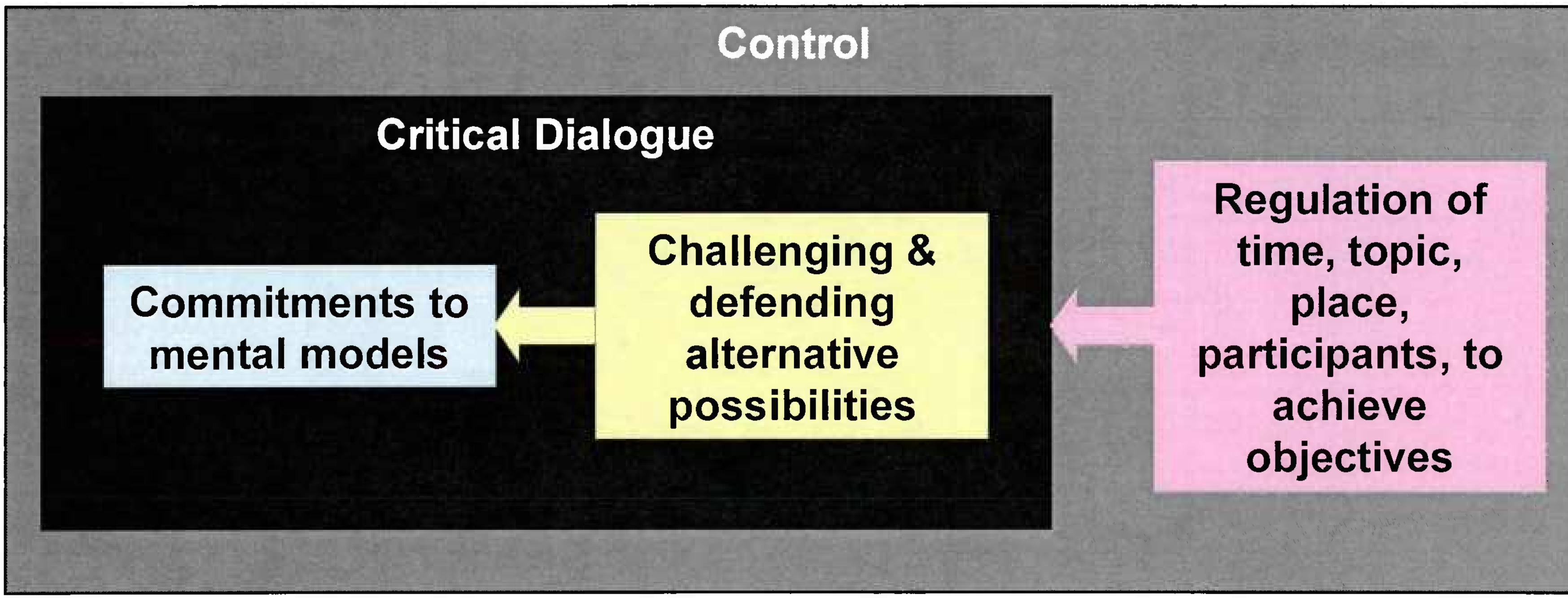
3. Does NDM offer any normative or prescriptive insight other than *"This is what experts do"*?

- What if there are no experts in a particular domain, task, or situation?
- Are these cases necessarily "non-naturalistic" or of lesser interest?
- Does NDM have anything distinctive to say about deliberative reasoning?

Normative Blend of Convention plus Functionality

DIALOGUE MODELS

Critical Thinking = Interplay of Three Perspectives



First person point of view: **Proponent**

Second person point of view: **Critic**

Third person point of view: **Facilitator - Judge**

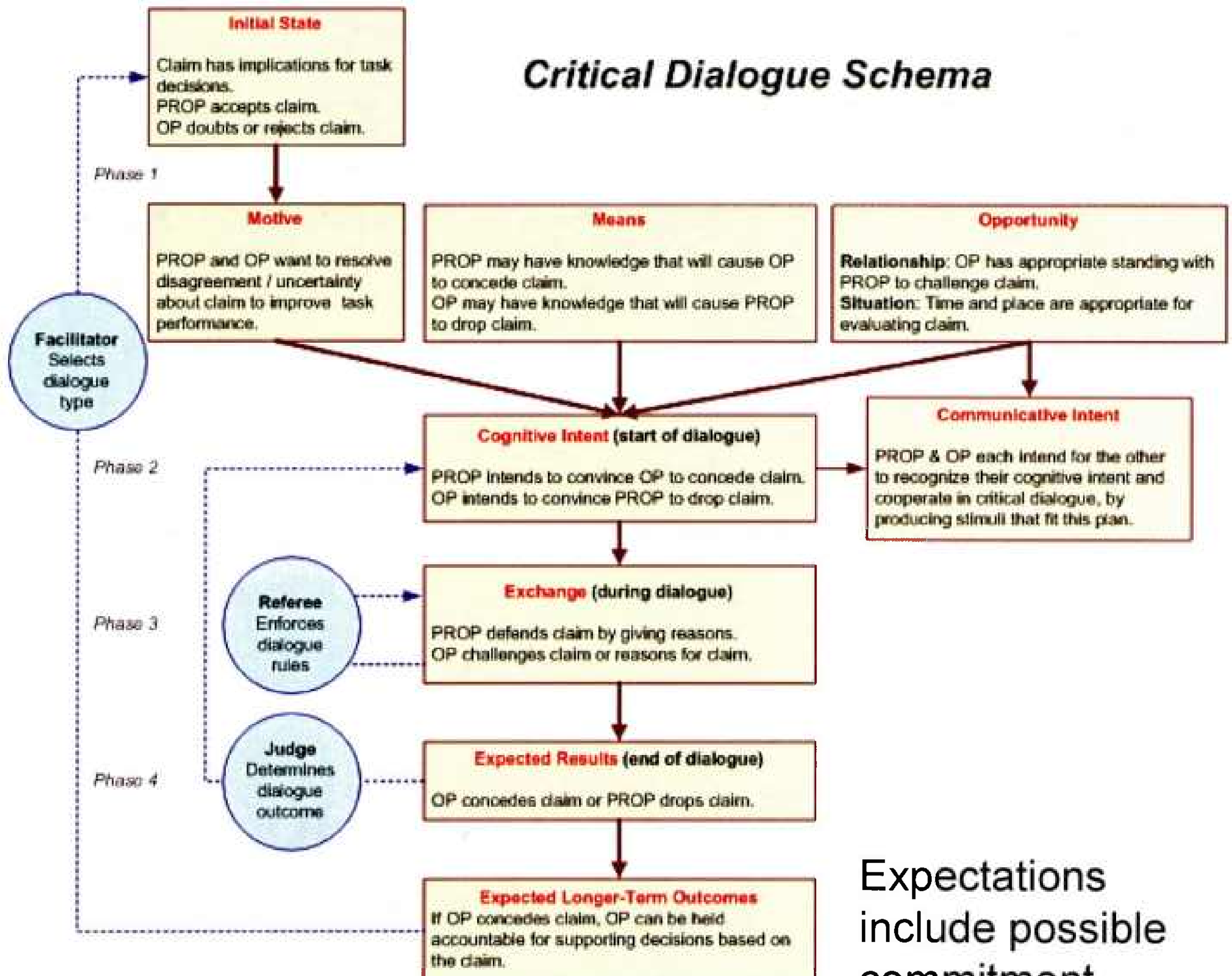
Critical Dialogue Involves Shared Intention and Expectations = Implicit Commitment

1. Recognize Disagreement

2. Commit to dialogue type, roles, topics

3. Exchange challenges and defenses

4. Resolve issue, change commitments accordingly



Expectations include possible commitment change

Example

1. One party **initiates** by **challenging** a view or plan accepted by the other.

2. The other party **recognizes** the intent to challenge, and **accepts** the role of **defender**.

3. Discussion continues the exchange of **challenges and answers**.

Example

1. One party initiates by challenging a view or plan accepted by the other.

2. The other party recognizes the intent to challenge, and accepts the role of defender.

3. Discussion continues the exchange of challenges and answers.

4. Dialogue ends either when time runs out, challenger drops doubt, or defender gives up claim.

Some Normative Implications for Exchange of Challenges & Defenses

- Once dialogue type and roles are accepted:
  - Moves are constrained by:
    1. Decision mode
    2. Dialogue schema
    3. Argumentation schemes
    4. Evolving mental models of the situation & plan
- Not just anything can be *appropriately* questioned in a *given* context!
- But almost anything can be appropriately questioned in *some* context (no privileged class of “data”)

Further Implications

Once implicit commitments are made to critical dialogue type, roles, topics, these issues should not intrude into exchange of defenses and challenges:

- No personal attacks, bargaining, appeals to authority, etc.
- Keep reasons relevant
  - A claim that is not challenged is conceded.
  - If a claim is challenged, it *must* be defended or dropped.
  - A claim cannot be defended *unless* it is challenged

Example

1. One party **initiates** by **challenging** a view or plan accepted by the other.

2. The other party **recognizes** the intent to challenge, and **accepts** the role of **defender**.

3. Discussion continues the exchange of **challenges and answers**.

Violates expectations that have developed. Discourages future constructive interaction.

Example

1. One party **initiates** by **challenging** a view or plan accepted by the other.

2. The other party **recognizes** the intent to challenge, but DECLINES invitation to play the role of **defender**.

LT, there may be some risk, but we can't make any changes now.

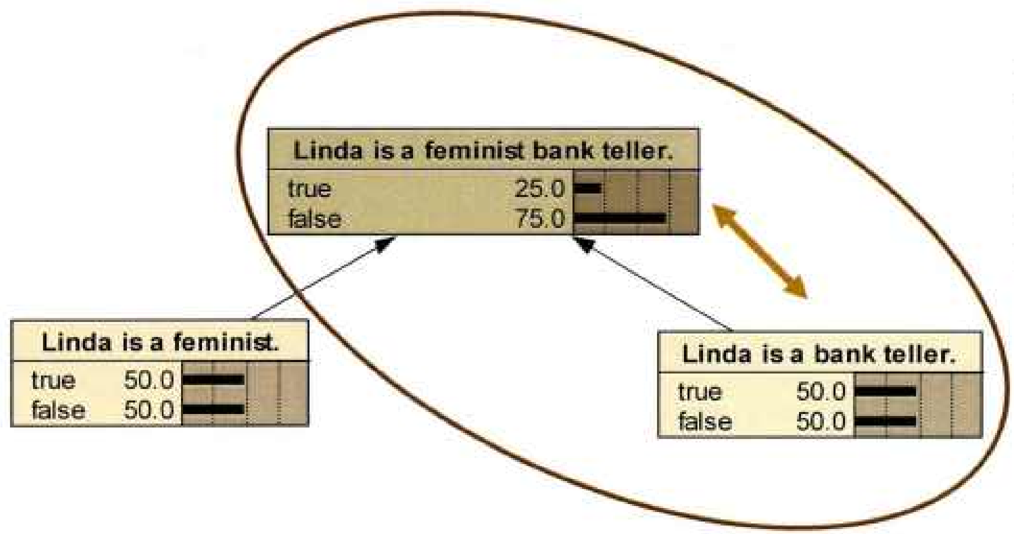
Can we count on the enemy crossing the river at that point?

Requester understands that in different circumstances, invitation might be accepted.

Normative Mistakes

IGNORING DIALOGUE CONTEXT: BIASES & HEURISTICS RESEARCH

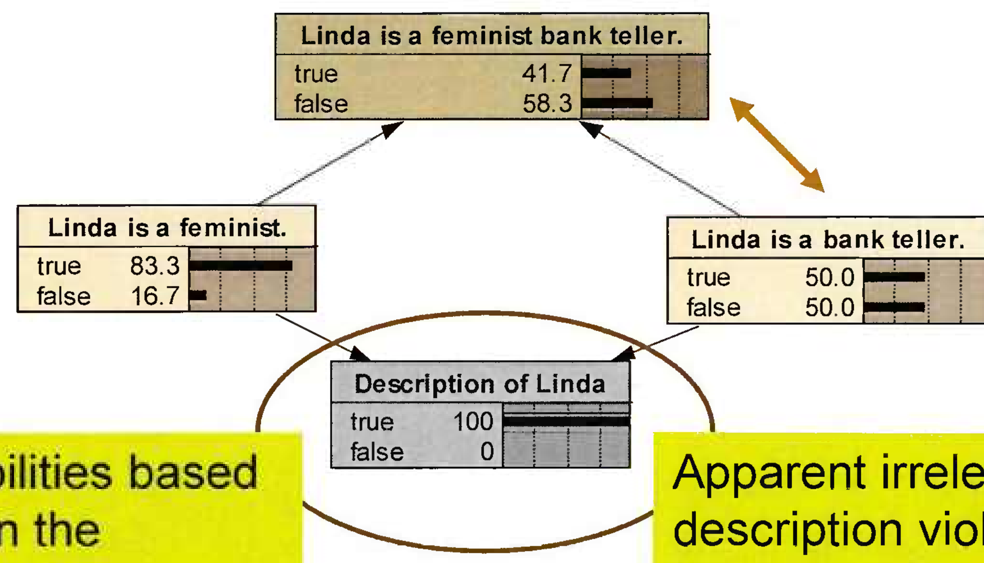
Conjunction "Bias"



Pure Logic??

Probability of two propositions is less than either one alone.

Breaking a Dialogue Rule



Pure Logic??

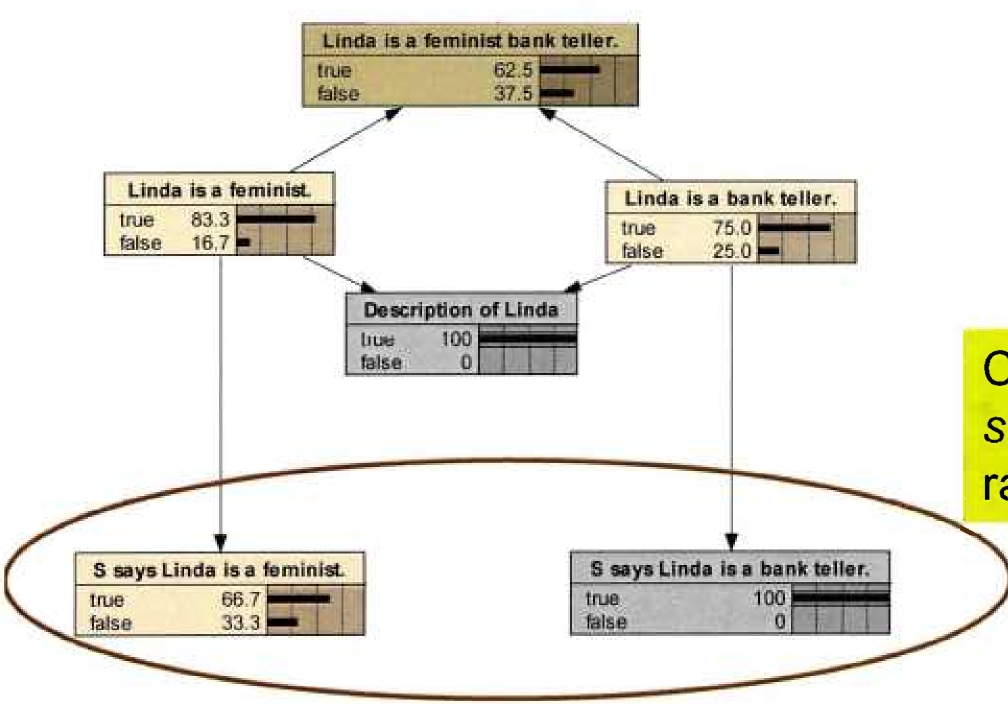
Probability of two propositions is less than either one alone.

Actual probabilities based on evidence in the description do not matter!

Apparent irrelevance of description violates norms of conversation.

Could truth of both propositions (conjunction) make a more "coherent" story *given the description*?

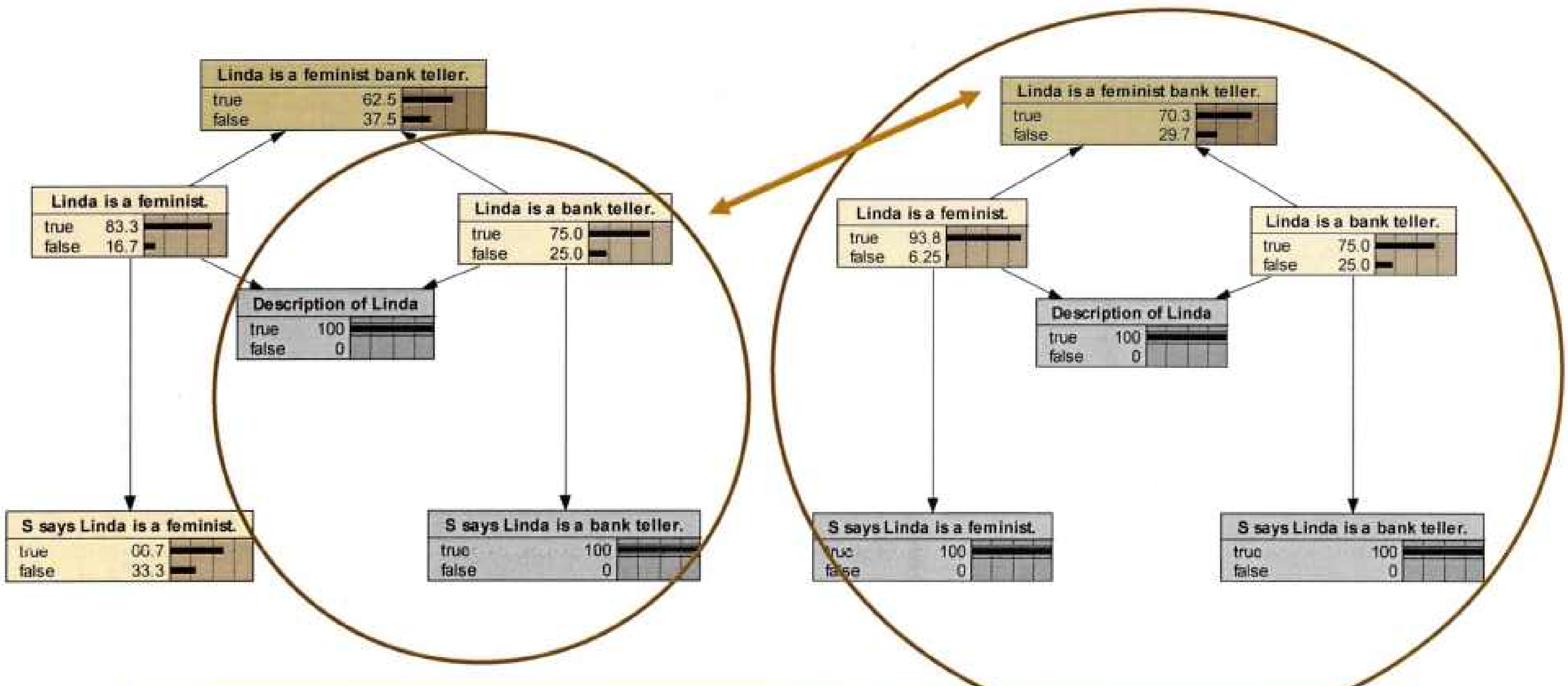
A Natural Story



Consider the "data" to be *statements* by a speaker *S*, rather than abstract propositions

Compare Two Stories

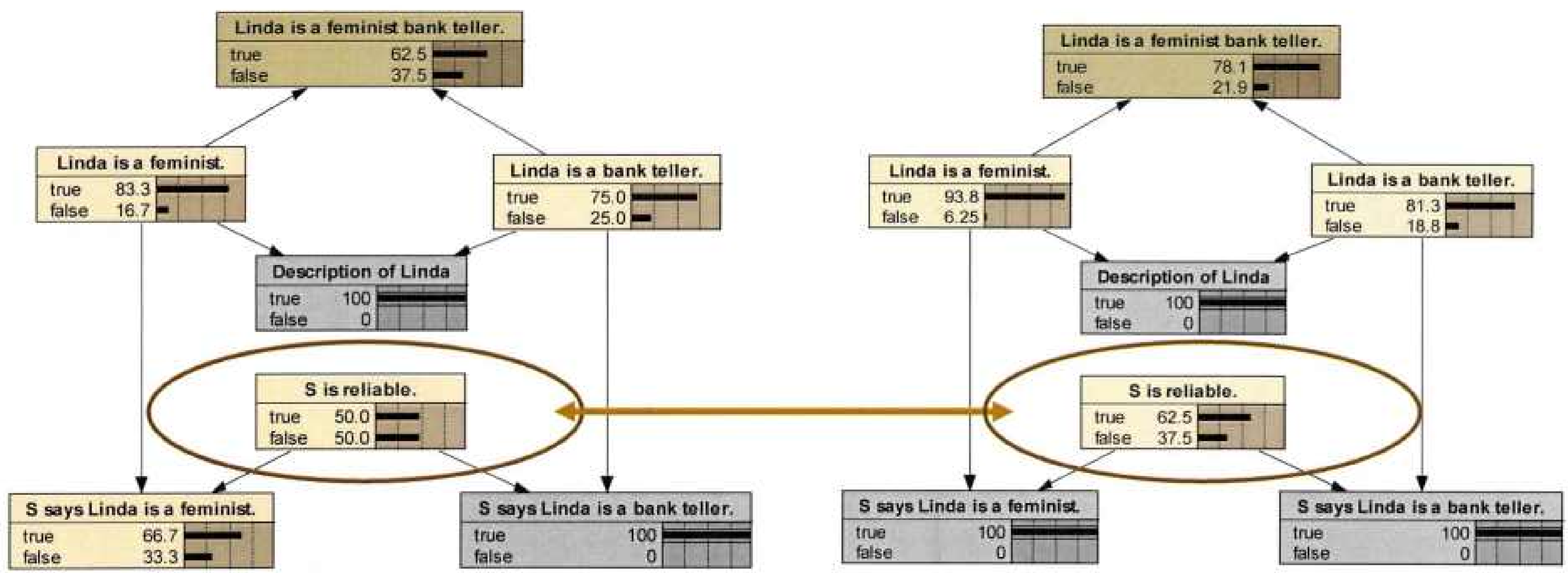
Now compare two *situations*, involving different claims by *S*.



No longer logically necessary for single claim to be more probable than conjunction: conditioned on different evidential situations!

Challenge Credibility

It is appropriate to question the credibility of *S* in the two situations, so add it as a factor.

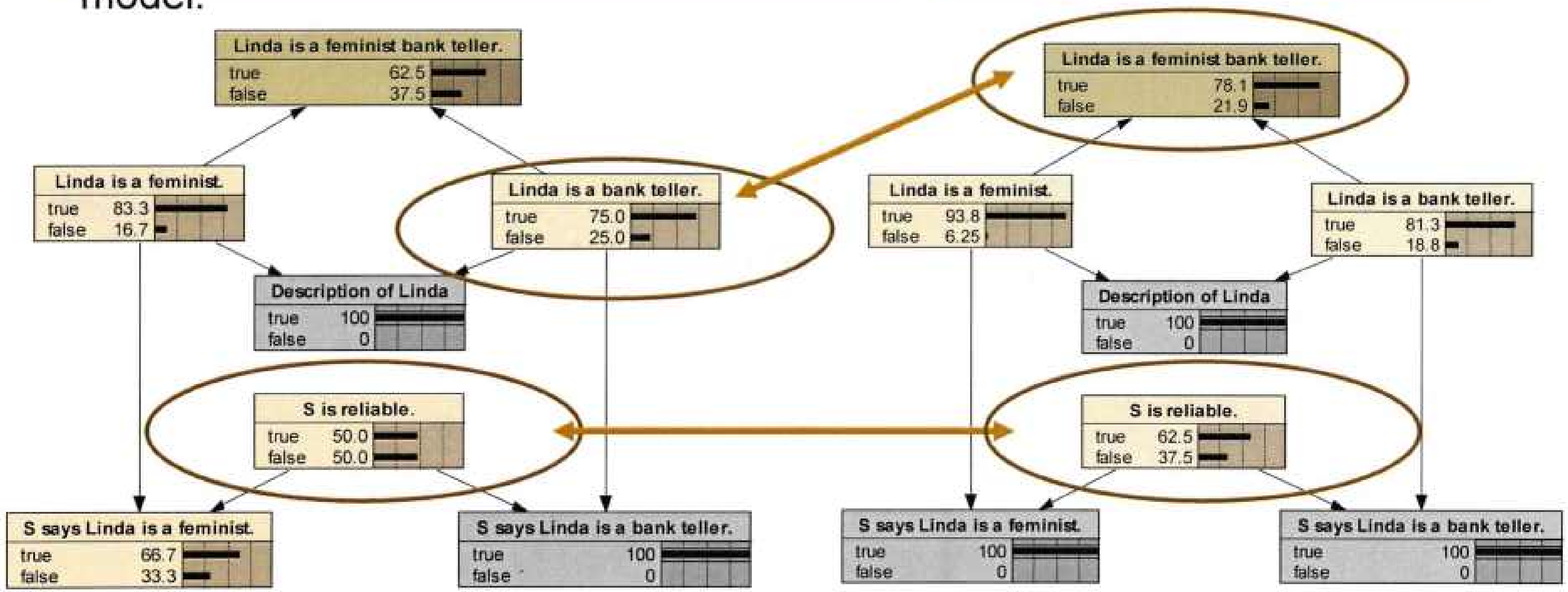


The credibility of *S* rises when S says that Linda is a feminist, because it fits the description.

Conjunction "Bias" Disappears

It is appropriate to question the credibility of *S* in the two situations, so add it to the model.

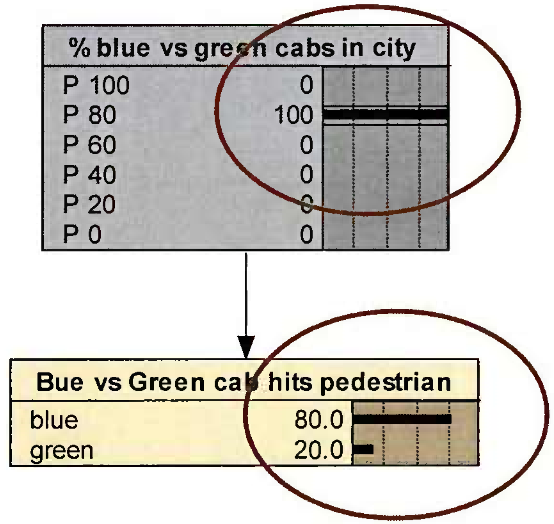
This offsets the loss of probability due to making two claims!



The credibility of *S* rises when S says that Linda is a feminist, because it fits the description.

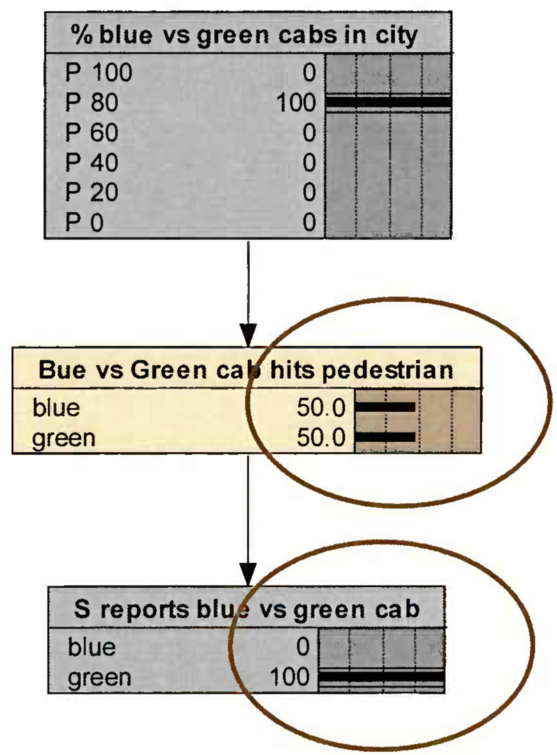
Base Rate "Bias"

With no other evidence, probability of guilt is determined by base rate of blue cabs, 80%.



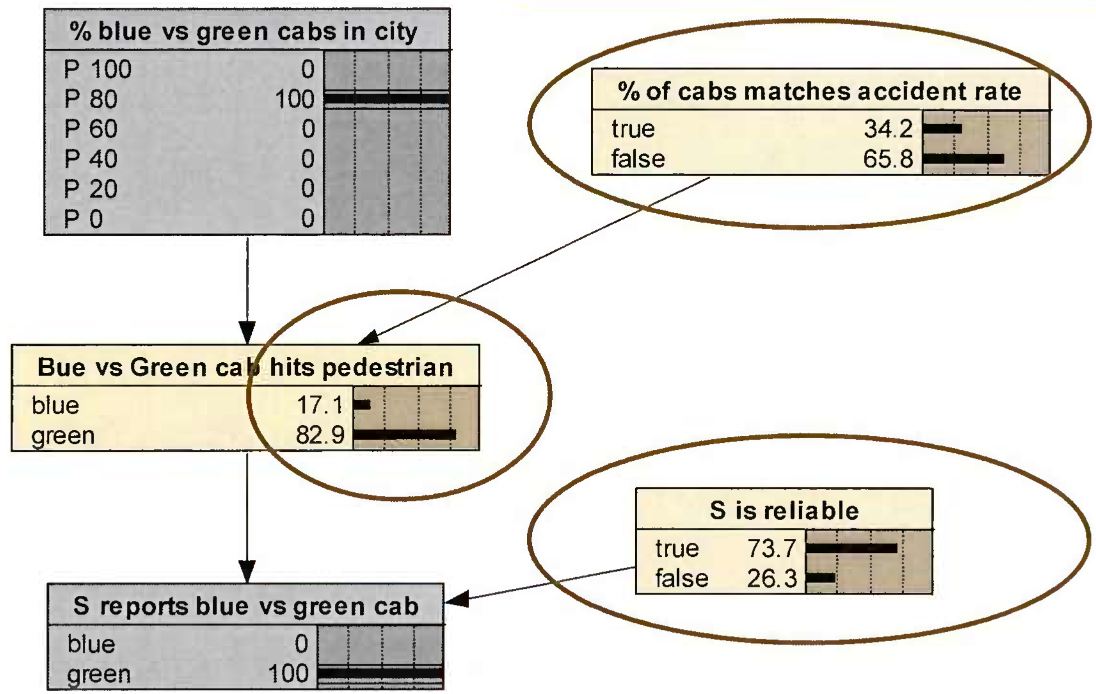
Base Rate "Bias"

Witness' conflicting testimony has reliability equal to base rate (80%) – so it brings probability to 50-50.



Base Rate "Bias"

Base rate and witness conflict. So, reliability of each can be appropriately challenged.



Initial trust in sources declines. If trust in witness (80%) was greater than trust in base rate (50%), result favors witness.

Appears to be ignoring base rate, but actually factoring in its credibility.

Dialogue Explanation

- Regard experimental inputs not as abstract propositions but as statements by a person
- Challenge to reliability is appropriate in the presence of conflict
- These challenges introduce credibility of source into deliberations

- But note: Critiquing is incremental and contextual. Decision makers *cannot* construct detailed models that somehow make *all* their assumptions explicit.

“Biases” Due to Matching

- Are self-control, altruism, and voting rational?
- Classical normative choice model (Expected Utility) says *No.*

- They involve **Matching**: determining appropriate or obligatory behavior by matching the situation to one’s role in a practice.
- Matching constrains the options available for commitment.

- Dialogues are idealized examples of practices with roles and rules.

“Biases” Due to Reassessment

- Existing commitments shape (and therefore bias) later decisions and increase resistance to reassessment.
- Path dependence is implied: Different decision makers will arrive at different conclusions due to differences in initial commitments.
  - Not a feature of classical normative theory.

Explanation in terms of Commitment

- Rejection of a *central* commitment demands large leap into an unmapped territory with low chance of success
  - Relevant factors & information sources are unknown or uncalibrated
  - Many plans and assessments are uprooted
- Better results are likely with critiquing and incremental modification of current commitments

Example: Confirmation Bias

- Wason rule task: Given a sequence (e.g., 2 4 6), figure out what the rule is by proposing new sequences and getting *yes/no* feedback.
- People consider hypotheses such as “even numbers” and produce positive examples (e.g., 10 20 26).
- Actual rule was “increasing values.”
- Widely cited conclusion: Confirmation bias = Unwillingness to seek evidence that might disconfirm a hypothesis.

JUST PLAIN WRONG: “No” would disconfirm the hypothesis.
Please NEVER CITE THIS EXPERIMENT UNCRITICALLY!!

Wason's Bad Logic

- Either of two strategies can disconfirm the hypothesis ("even numbers"):
  - Positive testing strategy (e.g., 10 20 26):

    Expected answer is *yes*, disconfirming answer is *no*.
  - Negative testing strategy (e.g., 5 7 11):

    Expected answer is *no*, disconfirming answer is *yes*.
- Note: You cannot *expect* disconfirmation of a hypothesis if you believe it to be true.
- So, why do participants prefer positive testing strategy??

Two-Part Explanation

DIALOGUE

- Conversational norms of relevance suggest that salient features of the positive instance (2 4 6) will be part of the correct rule.
- This is the initial commitment that guides subsequent guesses.
- Exploiting this evidence results in a tractable search strategy:
  - If answer is *no* (e.g., to 10 20 26), add another salient feature of the positive instance that is absent in the negative example (e.g. *consecutive* & even, or *less than 10* & even).
- Strategy is to search in vicinity of the hint assumed to have been given by experimenter.

Two-Part Explanation

STRUCTURE

- The experimenter violated these norms: Example was deceptive. Even if participant knows about deception, problem is unstructured, hence, intractable!
- Infinitely many possible rules fit any finite set of positive &/or negative instances. Participants need plausible assumptions to narrow the space!!
  - There is no efficient strategy for choosing negative instances or reacting to the answer. (No idea which of infinitely many features are responsible for *Yes* or *No* answers.)

Summary of Themes

1. Decision making can be defined comprehensively, consistently, and (I think) insightfully as *Commitment change*
2. Modeling in NDM highlights issues of dynamic knowledge representation and self-regulation
3. Normative implications emerge from blend of convention and functionality, which dialogue theory attempts to capture

A Mantra for Applications as well as Theory

Pattern recognition first, then verify and modify if time and costs allow!

- (1) NDM must develop pre-packaged templates, implemented in software, that offer general types of solutions to recurring problems in naturalistic domains
  - *Based on formal models and normative principles (e.g., mental models, decision modes, dialogue rules)*
  - *Apply to full range of processes that can change commitment*
- (2) NDM can support more particularized analysis to elaborate and modify the templates when time, stakes, and initial results warrant the investment.

- But that's another talk!

Forecasting Advancements In Weather-Related NDM Research

Daphne S. Ladue, Ph.D.

University of Oklahoma's Center For Analysis And Prediction Of Storms

Weather ... climate ... climate change ...

Weather has normals and extremes

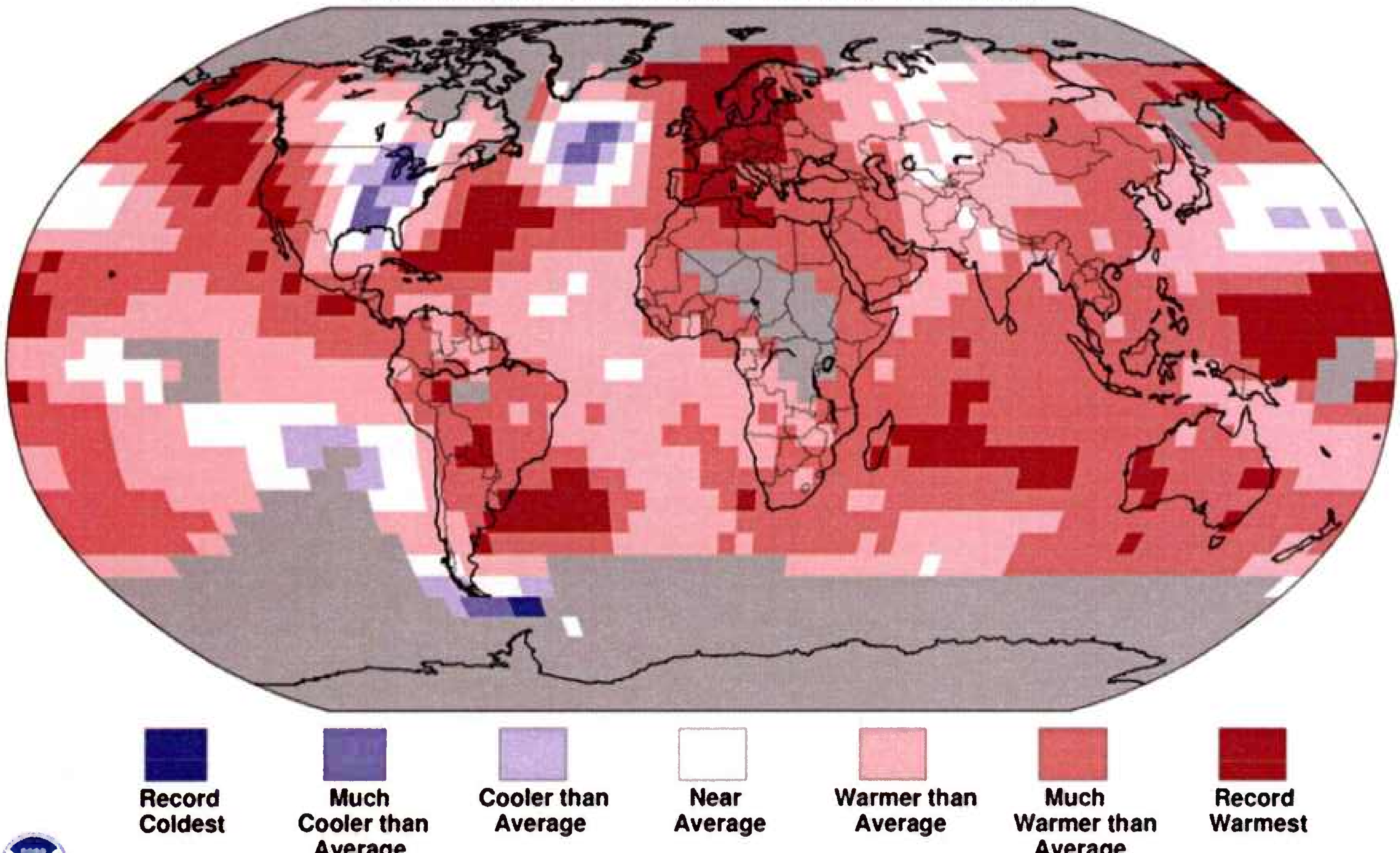
Climate is the mean, gross aspects of the system

“Climate is what you expect. Weather is what you get.”

Land & Ocean Temperature Percentiles Jan–Dec 2014

NOAA's National Climatic Data Center

Data Source: GHCN–M version 3.2.2 & ERSST version 3b

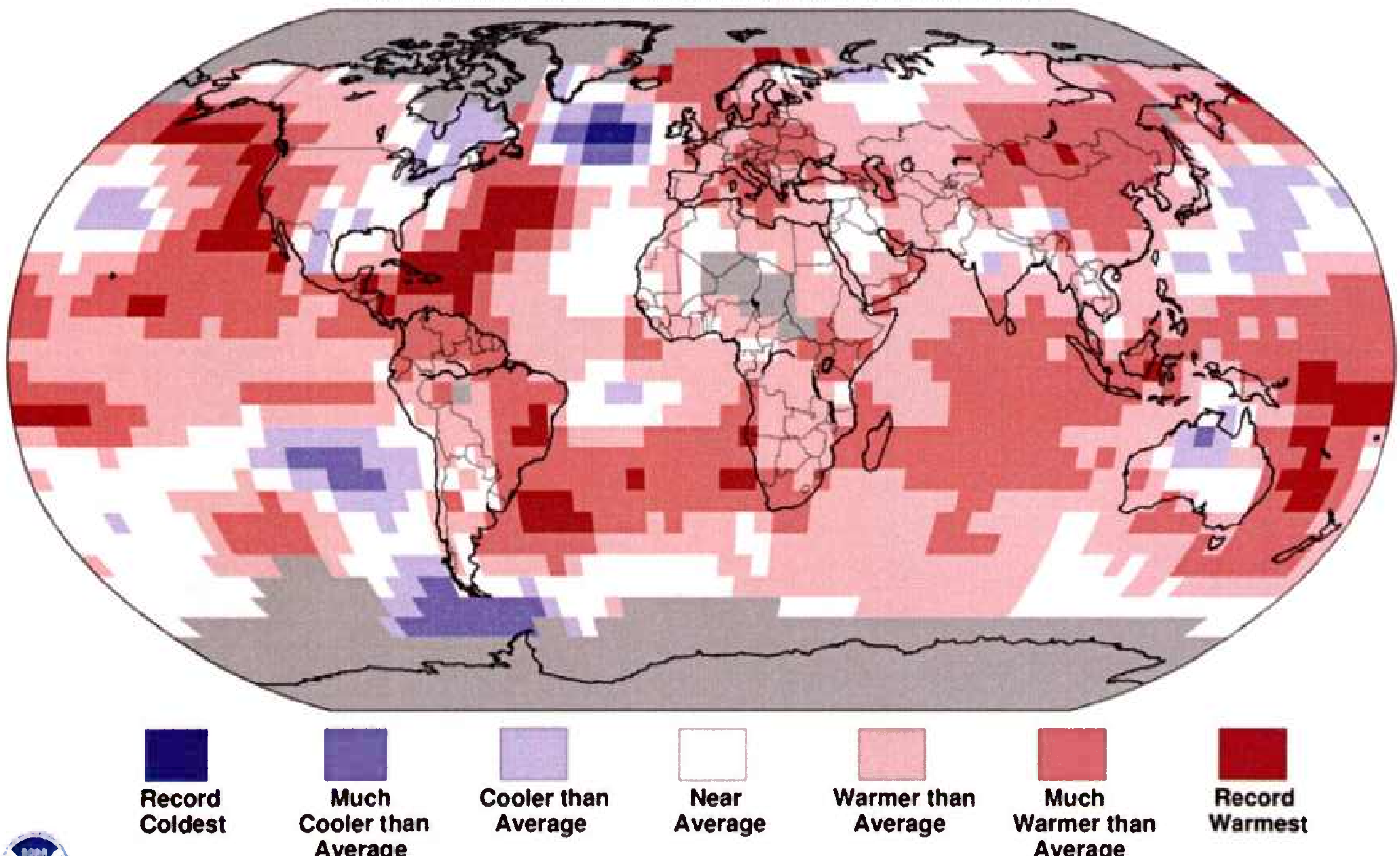


Mon Jan 12 19:34:46 EST 2015

Land & Ocean Temperature Percentiles Jan 2015

NOAA's National Climatic Data Center

Data Source: GHCN-M version 3.2.2 & ERSST version 3b

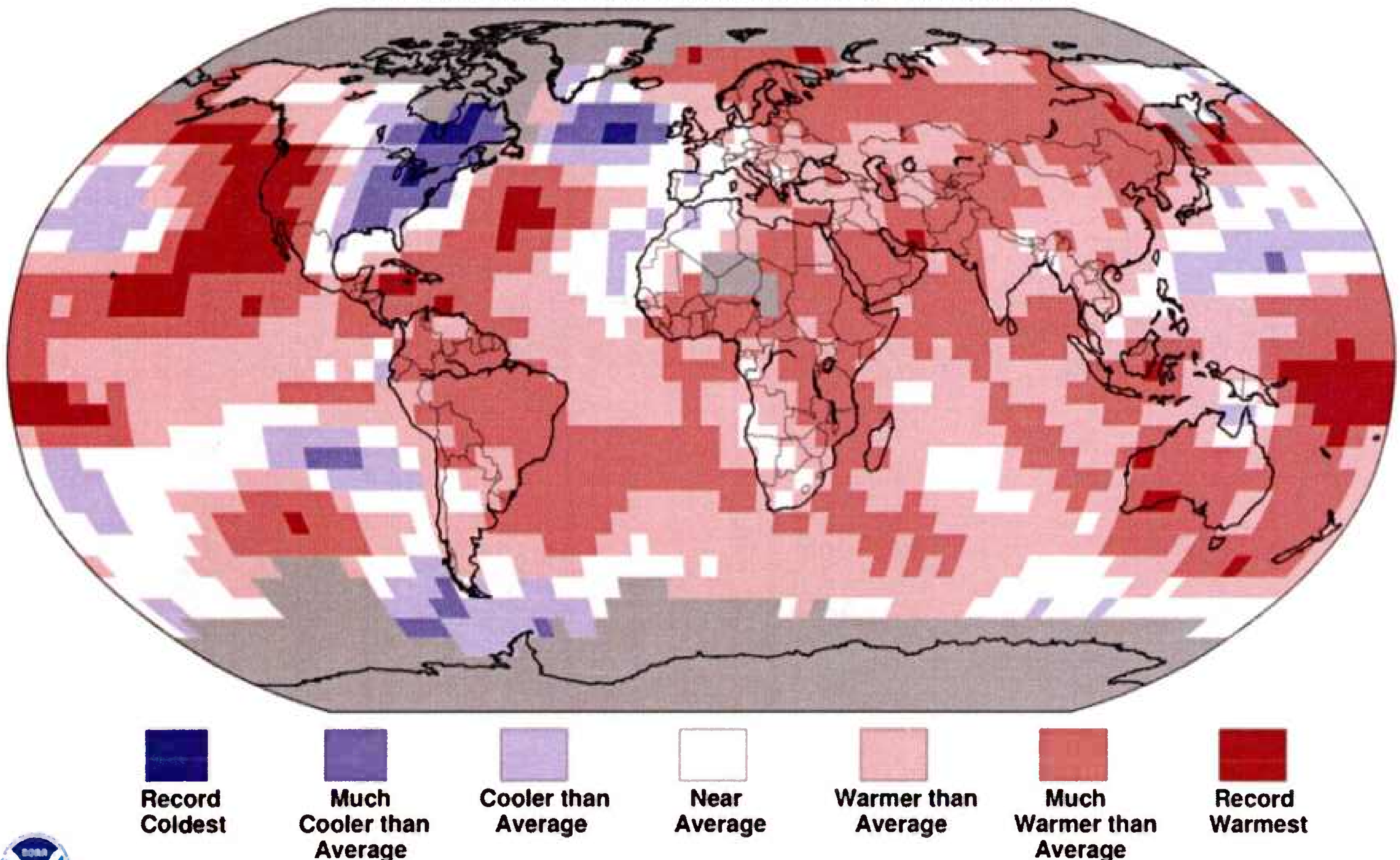


Mon Feb 16 19:13:29 EST 2015

Land & Ocean Temperature Percentiles Feb 2015

NOAA's National Climatic Data Center

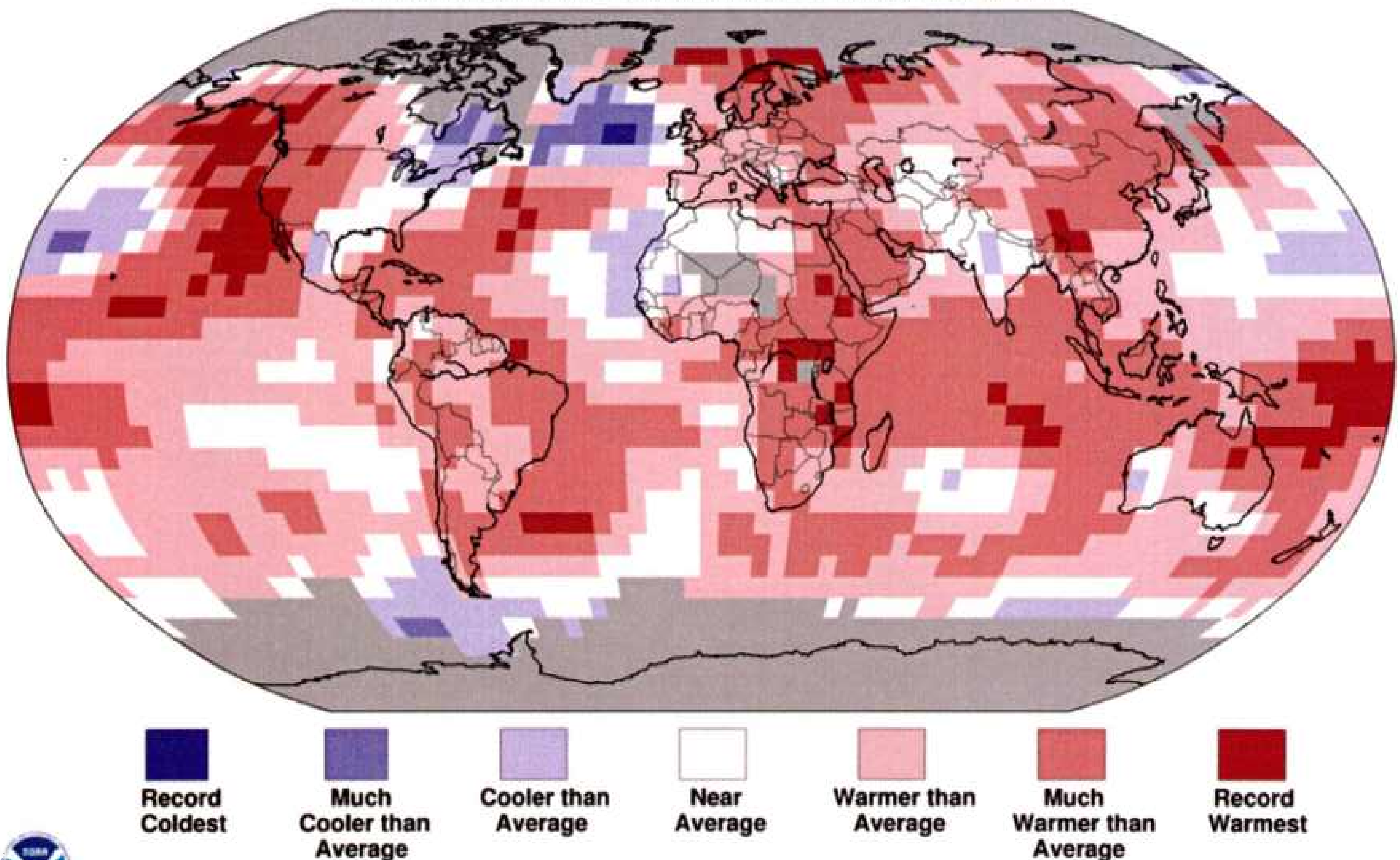
Data Source: GHCN–M version 3.2.2 & ERSST version 3b



Land & Ocean Temperature Percentiles Mar 2015

NOAA's National Climatic Data Center

Data Source: GHCN–M version 3.2.2 & ERSST version 3b



Tue Apr 14 12:31:43 EDT 2015

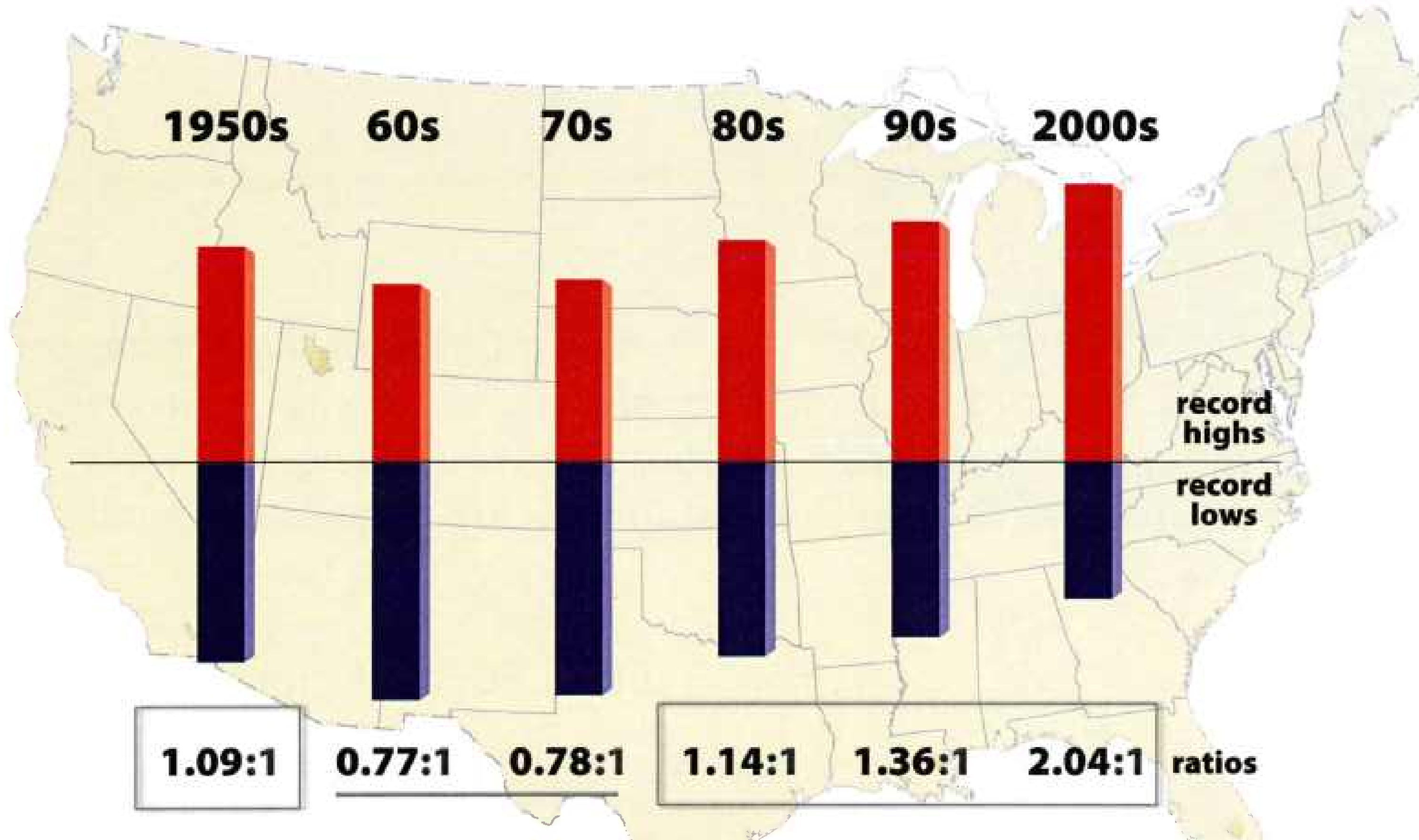
BELIEF?

I've been troubled by the whole notion of whether someone "believes" in climate change.

Science isn't about belief.

And it isn't about proving things! http://undsci.berkeley.edu

Science is discovery. The result is an maturing, sometimes evolving understanding, as scientists poke holes and test their ideas. They come at subjects from multiple angles to try to understand it thoroughly.

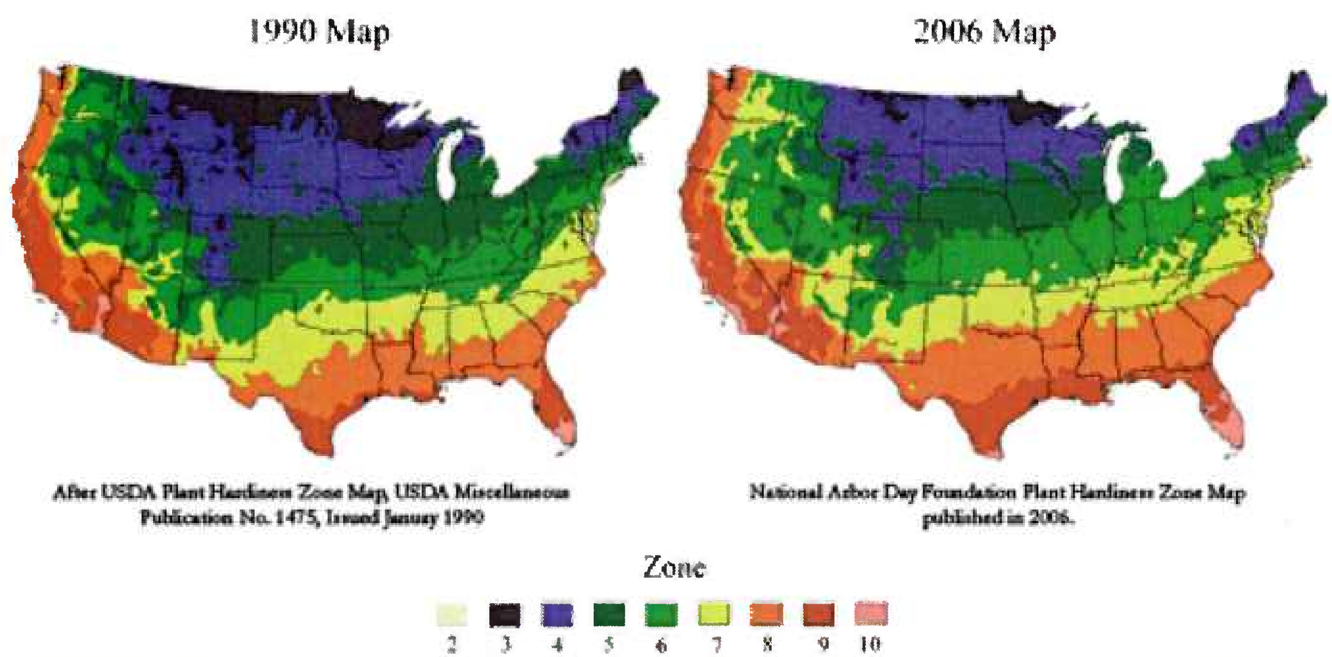


With climate change we still get cold days, and set cold temperature records. That's what *weather* does — vary a lot! — while *climate* is the background trend underneath.

GARDENING

Know a gardener?

The USDA confirmed what gardeners already knew with a new hardiness map for the U.S.

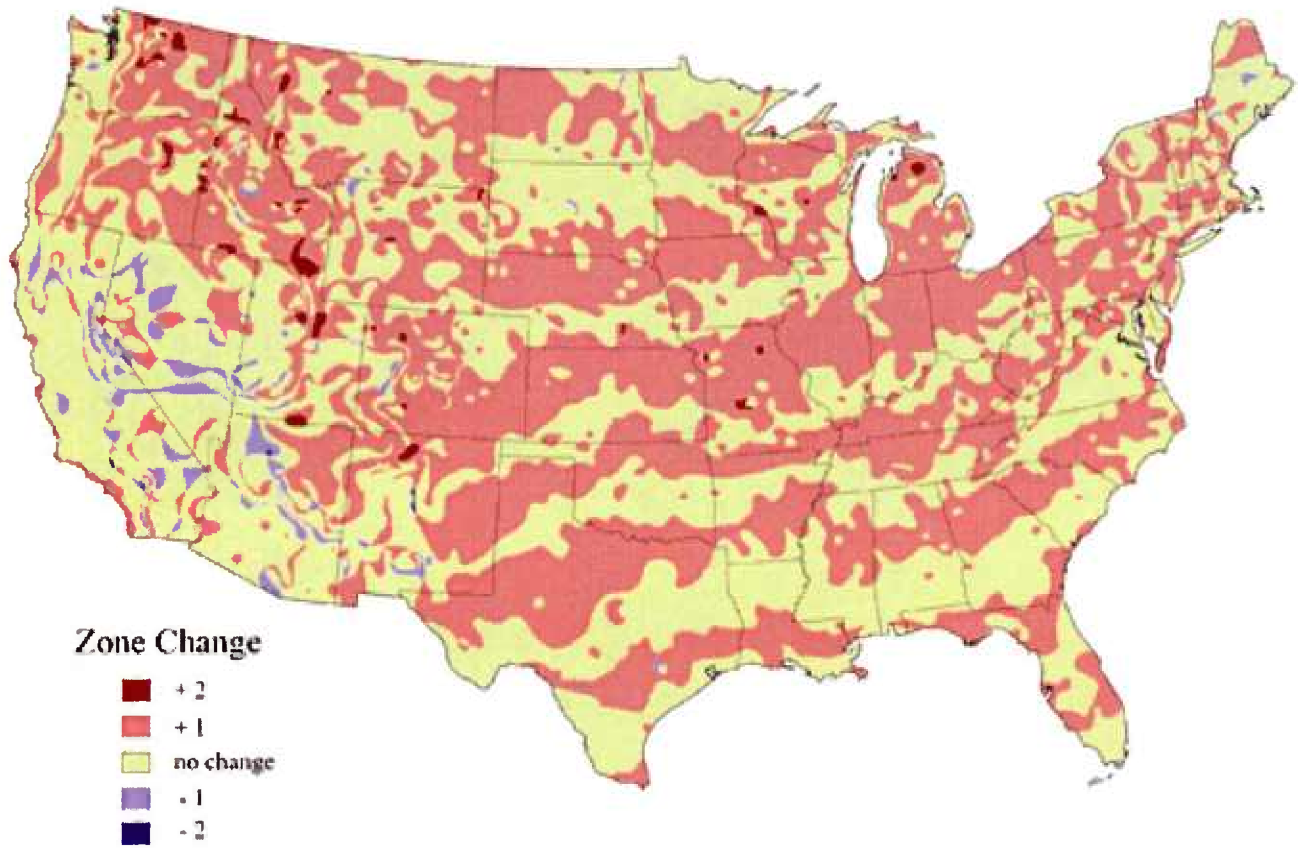


After USDA Plant Hardiness Zone Map, USDA Miscellaneous Publication No. 1475, Issued January 1990

National Arbor Day Foundation Plant Hardiness Zone Map published in 2006.

USDA Hardiness Zones are based on average annual minimum winter temperatures, which have been rising!
http://planthardiness.ars.usda.gov/PHZMWeb/

Differences between 1990 USDA hardiness zones and 2006 arborday.org hardiness zones reflect warmer climate



http://www.arborday.org/media/map_change.cfm

BIRD MIGRATION

Birders are rather enthusiastic about being the first to see things — and skeptical about others' claims (so they all flock to the area to see for themselves). Their "citizen science" records are pretty good.

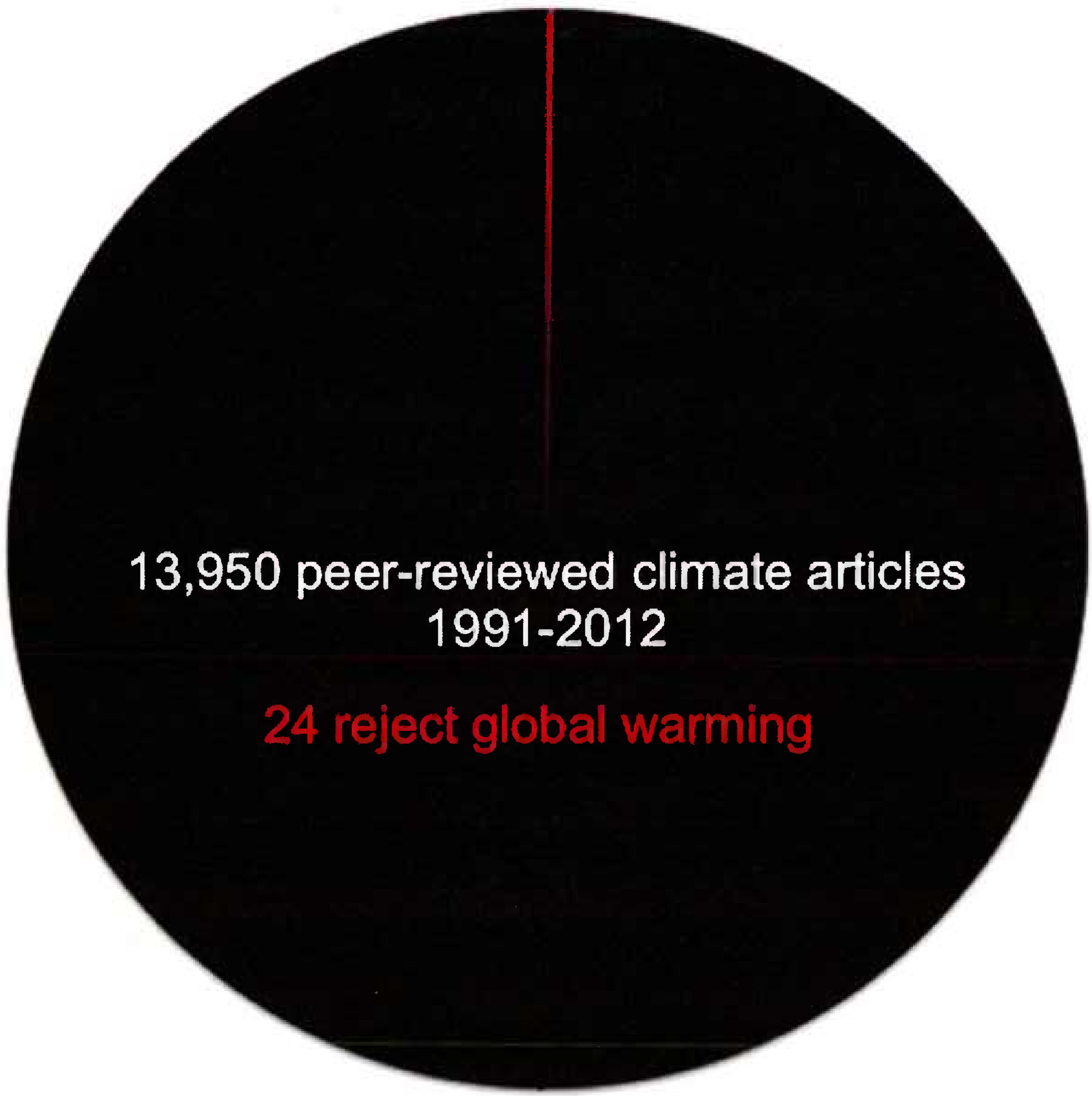
http://birdsandclimate.audubon.org

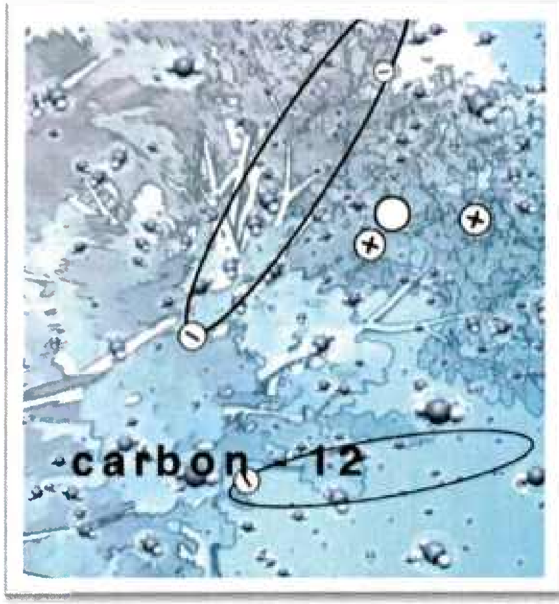
http://www.nytimes.com/2013/11/19/science/an-earlier-exodus-amid-climate-change.html

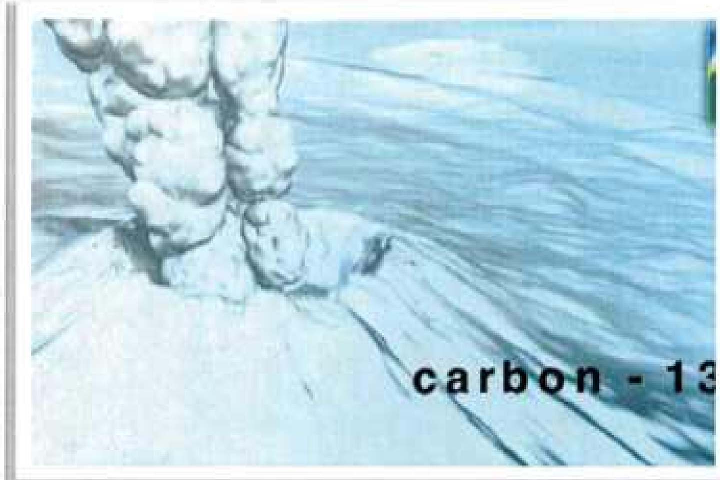
CAN'T STAND THE HEAT

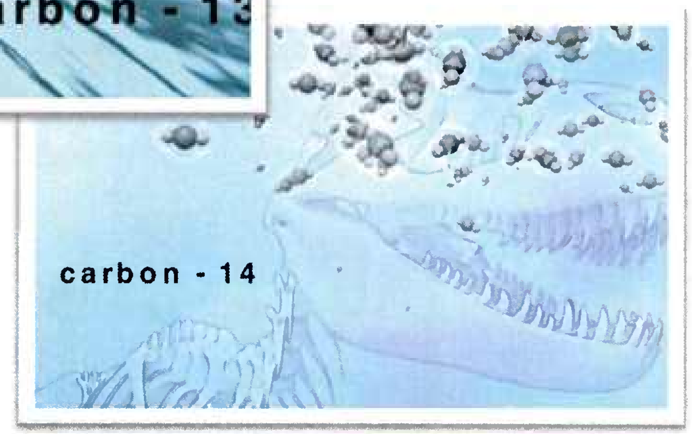
Winter 2014-2015 Temperature Percentiles Record Warmest Record Coldest

Adapted from NOAA National Climatic Data Center, State of the Climate Global Analysis for February 2015, published online March 2015. http://www.ncdc.noaa.gov/sotc/global


13,950 peer-reviewed climate articles
1991-2012
24 reject global warming





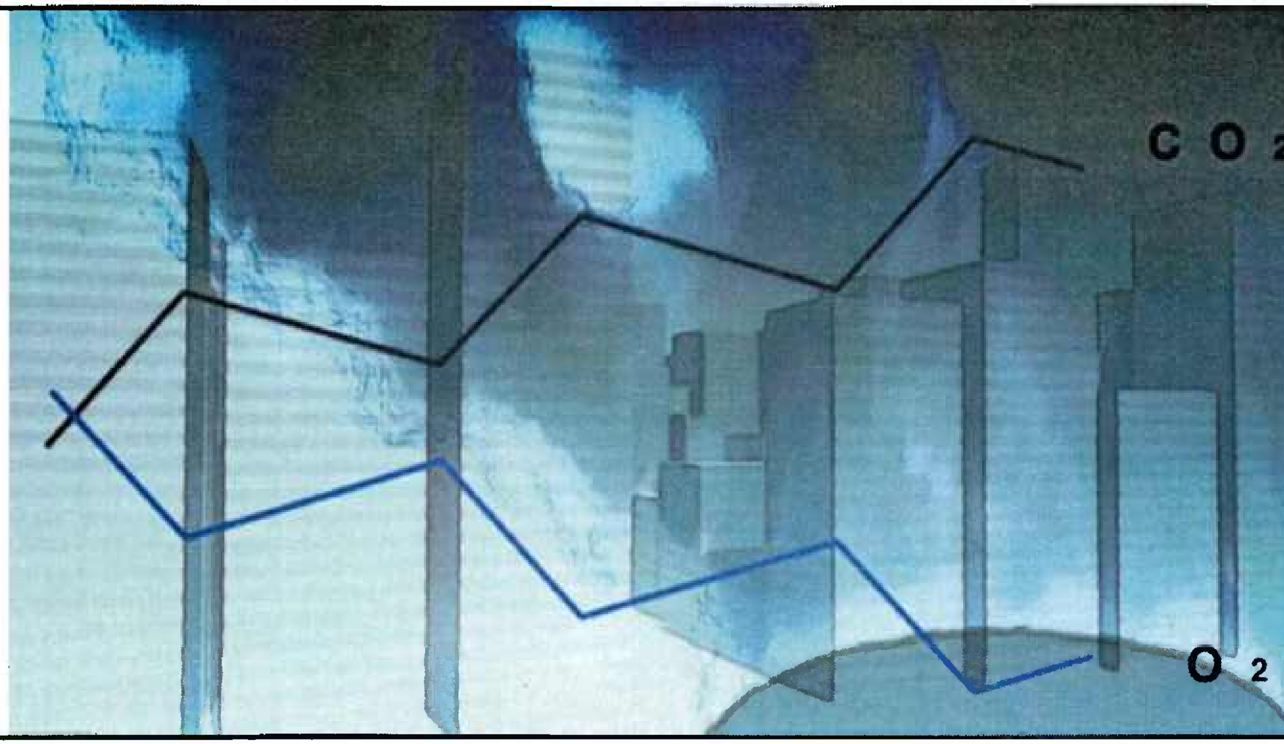


C_{12}: mostly from plants; 98.9% of all carbon on Earth

C_{13}: mostly from volcanoes

C_{14}: from recently-living plants; quickly decays to C_{12}

We also see that the rise in CO_2 goes with a fall of O_2, meaning the excess CO_2 is from burning something.



The heart of climate change is clear: it is our burning of fossil fuels.

HOW DOES NDM APPLY?

NDM: how do people make tough decisions under difficult conditions…

- **limited time**
- **uncertainty**
- **high stakes**
- **unstable conditions**

“People use prior experience to rapidly categorize situations. …some kind of synthesis of their experience…”

“…a blend of intuition [pattern matching] and analysis [metal simulation].”

Klein 2008

ENVIRONMENT

Threat to Bottom Line Spurs Action on Climate

By CORAL DAVENPORT JAN. 23, 2014

WASHINGTON — Coca-Cola has always been more focused on its economic bottom line than on global warming, but when the company lost a lucrative operating license in India because of a serious water shortage there in 2004, things began to change.

A Coke bottling plant in Winona, Minn. The company has been affected by global droughts.
Andrew Link/Winona Daily News, via Associated Press

Today, after a decade of increasing damage to Coke's balance sheet as global droughts dried up the water needed to produce its soda, the company has embraced the idea of climate change as an economically disruptive force.

"Increased droughts, more unpredictable variability, 100-year floods every two years," said Jeffrey

Meteorologists Shape Fashion Trends

Marilynn K. Yee/The New York Times

Eliot Peyser, left, and Fredric Stollmack of Weatherproof, a clothing manufacturer. Weatherproof has taken out what amounts to a $10 million insurance policy against unusually warm weather.

By MICHAEL BARBARO
Published: December 2, 2007

In the capricious world of fashion, where hemlines, fabrics and colors fall in and out of favor with breathtaking speed, designers and retailers have always relied on one constant — the orderly changing of the seasons.

But now it seems the seasons have become as fickle as fashion.

NY Times
December 2, 2007

The $200B American apparel industry is hiring climatologists to help them anticipate seasonal clothing demand

Impact: Medium
Time frame: 0–3 years

Brazilian drought

Description:
The ongoing drought in Brazil represents a serious threat to the country. Since 2014, levels of rainfall are far below average levels, leading to water shortage and extremely low levels in the water reservoirs. Some fresh water reservoirs close to São Paulo reached levels below 5% of their maximum capacity in early 2015.[1] The problem is worsened by polluted rivers, deforestation and population growth.

Poor planning and politics further aggravate the situation, with critics alleging that state authorities have failed to respond quickly enough to the crisis. Furthermore, Brazil's water infrastructure suffers from serious underinvestment and mismanagement, with more than 30% of São Paulo's treated water lost due to leaks and pilfering.[2]

Since Brazil is highly dependent on hydroelectric power, the drought is also impacting energy generation and supplies. At least six cities have been hit by blackouts due to weak hydroelectricity generation and high demand for air conditioning as temperatures soar over 35°C. In response, utilities are burning more fossil fuels, adding to the cost of energy and greenhouse gas emissions. Without major rainfall, power rationing is expected for the south-eastern region by latest in May 2015.[3]

Brazil's agricultural sector is also affected by the ongoing drought. Production of Arabica coffee beans, for instance, already fell 15% in 2014, pushing up the commodity's global price by almost half.[4]

Impact: High
Time frame: 0–3 years

Super nat cats

Description:
Large natural catastrophes represent major threats to the economy and society and create significant losses. Two cases in point are atmospheric river events and volcanic eruptions, both of which are not yet sufficiently taken into account as serious disruptors by a wide range of stakeholders.

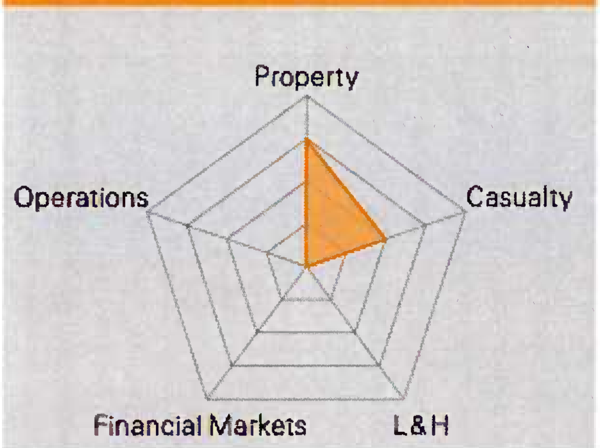
Atmospheric rivers (AR) are narrow corridors of concentrated moisture in the atmosphere. One of these corridors runs along the US West Coast. Between 1997 and 2006, California experienced more than 40 AR events that, while providing a valuable water supply for the region, also posed a threat in terms of extreme flash flooding and heavy snowfall. The US Geological Service published a study on a winter storm scenario called ARk, looking at the impact of an AR event with a return period of 1 000 years. Findings indicate that flooding would overwhelm flood protections in many areas, resulting in the evacuation of more than a million residents, direct property damage of nearly USD 400 billion and business interruption costs of about USD 325 billion. Insured property claims would add up to about USD 20–30 billion. These numbers are higher than those associated with an earthquake in southern California with similar likelihood.

[redacted]

Swiss Re

Swiss Re SONAR report: New emerging risk insights
May 2015

Impact: **Medium**
Time frame: **0–3 years**



Wildfires

Description:
According to Lloyd's, wildfires accounted for USD 7.9 billion in insured losses in the US between 2002 and 2011, an increase from USD 1.7 billion in the previous decade[9]. Up to 90% of all wildfires in the US are caused by humans, while the rest are started by lightning or lava.[10]

Wildfire losses materialise 52% in casualty covers, 40% in property and 9% in marine. There is also a marked increase in frequency and severity of these losses in the last decade: A review of losses at Swiss Re showed that globally 90% of insurance losses due to wildfires occurred since 2002. This is in line with projections from the Intergovernmental Panel on Climate Change (IPCC) which states in its Fifth Assessment Report that fire frequency is expected to increase with human-induced climate change, especially where precipitation remains the same or is reduced.[11] From a re/insurance perspective this implies that wildfires have a high potential to be underestimated with respect to frequency and severity in the coming years.

Munich RE

Climate Change

Climate Change is a subject that concerns us all

For more than 40 years, Munich Re has been dealing with climate change and the related risks and opportunities for the insurance industry. Our approach to coping with this challenge is holistic and based on the following pillars: risk assessment – insurance solutions – asset management.

Mission & Vision

Geoscientists at Munich Re have been analysing the effects of climate change on the insurance industry since the 1970s. Because of natural and anthropogenic changes to probability distributions for meteorological and hydrological parameters, a risk of change has resulted in the portfolios of Munich Re and its clients. You can find details of Munich Re's mission and vision here.

Based on our analyses over a period of more than 40 years, we have developed a comprehensive strategy whereby we identify and assess the risks and reflect them in our business processes. The physical changes caused by climate change and political/regulatory measures aimed at adapting to it and reducing emissions are helping to develop new topic areas that we assess in detail in terms of their relevance for our business activities (on both the asset and liability sides of the balance sheet).

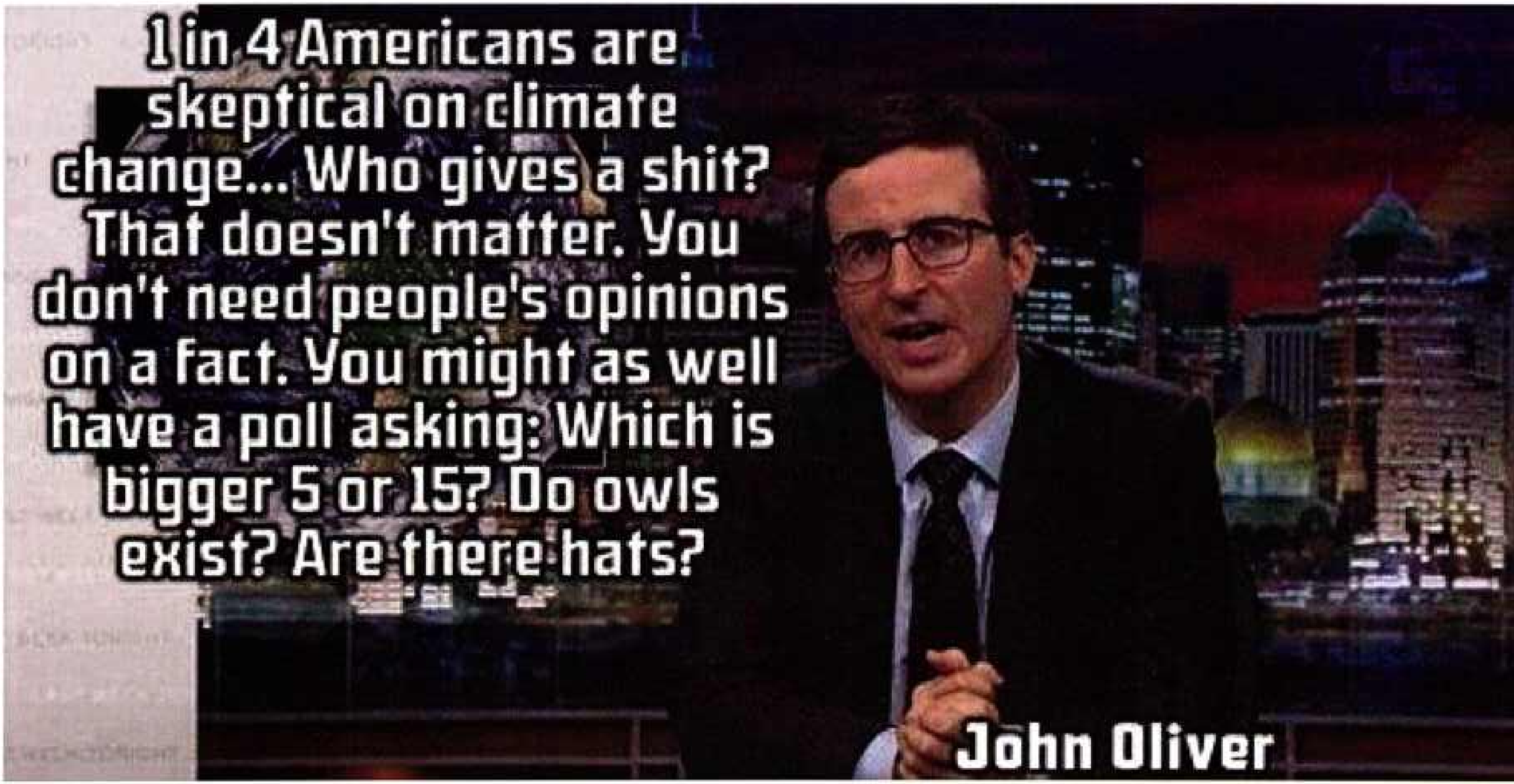
Together, the mission and the vision define the framework for Munich Re's climate strategy.

Our **mission** describes how we identify and address the changes resulting from climate change. It also underlines our resolve to treat the challenges and opportunities arising from climate change as a long-term, strategic topic.

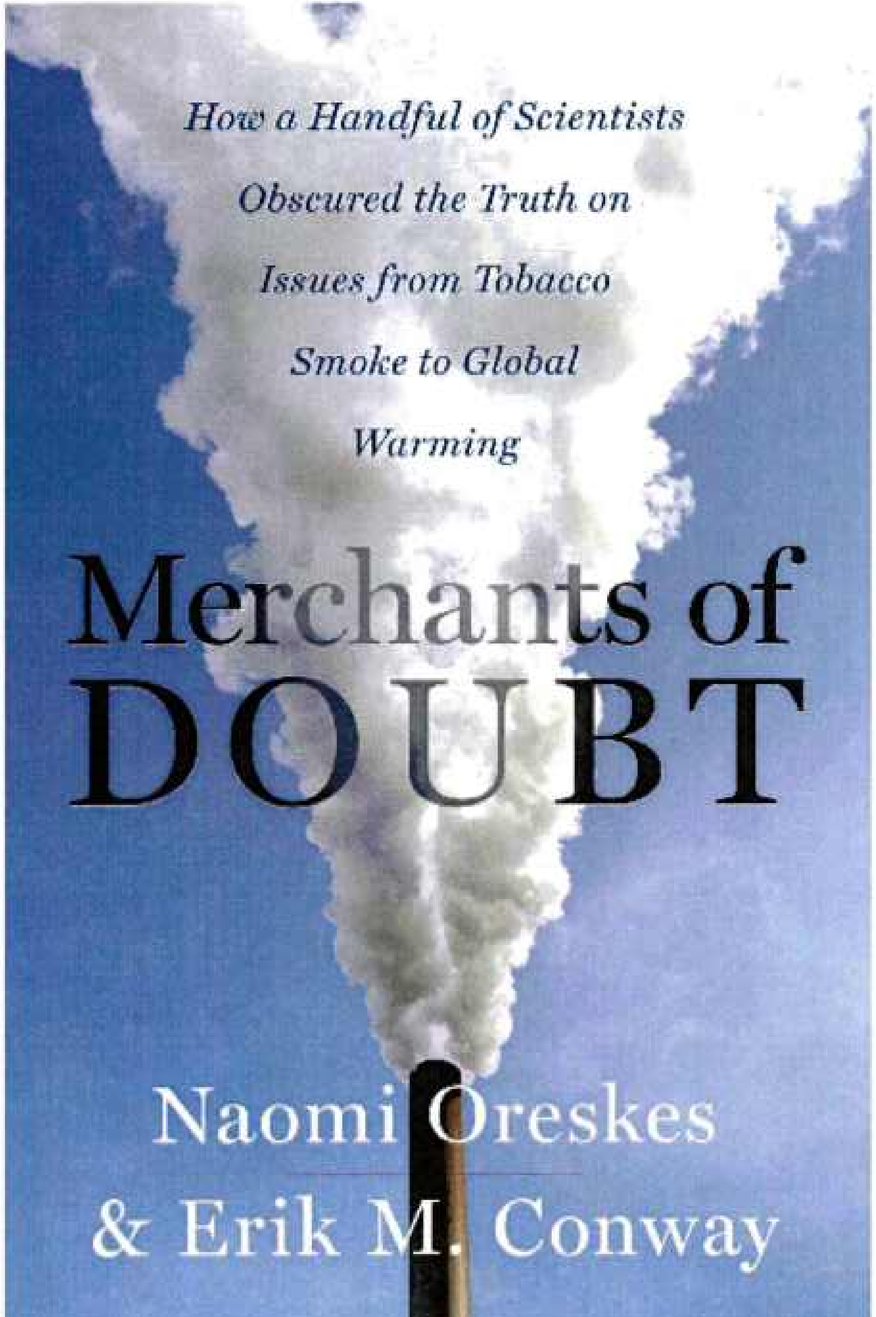
» Mission

The **vision** outlines our responses as a leading reinsurer to the challenges thrown up by climate change. We believe it is important to place a strong emphasis on managing climate change, rather than simply responding passively to it.

» Vision


1 in 4 Americans are skeptical on climate change... Who gives a shit? That doesn't matter. You don't need people's opinions on a fact. You might as well have a poll asking: Which is bigger 5 or 15? Do owls exist? Are there hats?
John Oliver

"Senator Snowball" by Andrew David Cox. ©2015
http://reverbpress.com/cartoon/editorial-cartoon-snowball-senator/



The idea is to delay action by sowing seeds of doubt.

“Keep the controversy alive.”

Not only are the tactics the same, the same *people* are involved.

Fred Singer, Fred Seitz, and a handful of other scientists joined forces with think tanks and private corporations to challenge scientific evidence on a host of contemporary issues. In the early years, much of the money for this effort came from the tobacco industry; in later years, it came from foundations, think tanks, and the fossil fuel industry.

Oreskes, Naomi; Conway, Erik M. (2010-06-03). Merchants of Doubt (p. 6). Bloomsbury Publishing Plc. Kindle Edition.

[Fred] Singer was a physicist— in fact, the proverbial rocket scientist — who became a leading figure in the development of Earth observation satellites, serving as the first director of the National Weather Satellite Service and later as chief scientist at the Department of Transportation in the Reagan administration.

Oreskes, Naomi; Conway, Erik M. (2010-06-03). Merchants of Doubt (p. 5). Bloomsbury Publishing Plc. Kindle Edition.

JUST ONE OF THE STRATEGIES:

"Little's committee prepared a booklet, A Scientific Perspective on the Cigarette Controversy, which was sent to 176,800 American doctors. Fifteen thousand additional copies were sent to editors, reporters, columnists, and members of Congress. A poll conducted two years later showed that "neither the press nor the public seems to be reacting with any noticeable fear or alarm to the recent attacks."

"...in 1957 the Tobacco Industry Research Committee published 350,000 copies of a new pamphlet, Smoking and Health, mostly sent to doctors and dentists.

Oreskes, Naomi; Conway, Erik M. (2010-06-03). Merchants of Doubt (p. 18, 20). Bloomsbury Publishing Plc. Kindle Edition.

The Heartland INSTITUTE

A nonprofit organization devoted to discovering, developing, and promoting free-market solutions to social and economic problems.

March 2013

Dear Professor of Environmental Studies:

The Heartland Institute asked me to send you the enclosed copy of *The Mad, Mad, Mad World of Climatism* with my strong endorsement. I am happy to do so. Also enclosed is a copy of an excellent 10-minute DVD titled *Unstoppable Solar Cycles: The Real Story of Greenland.*

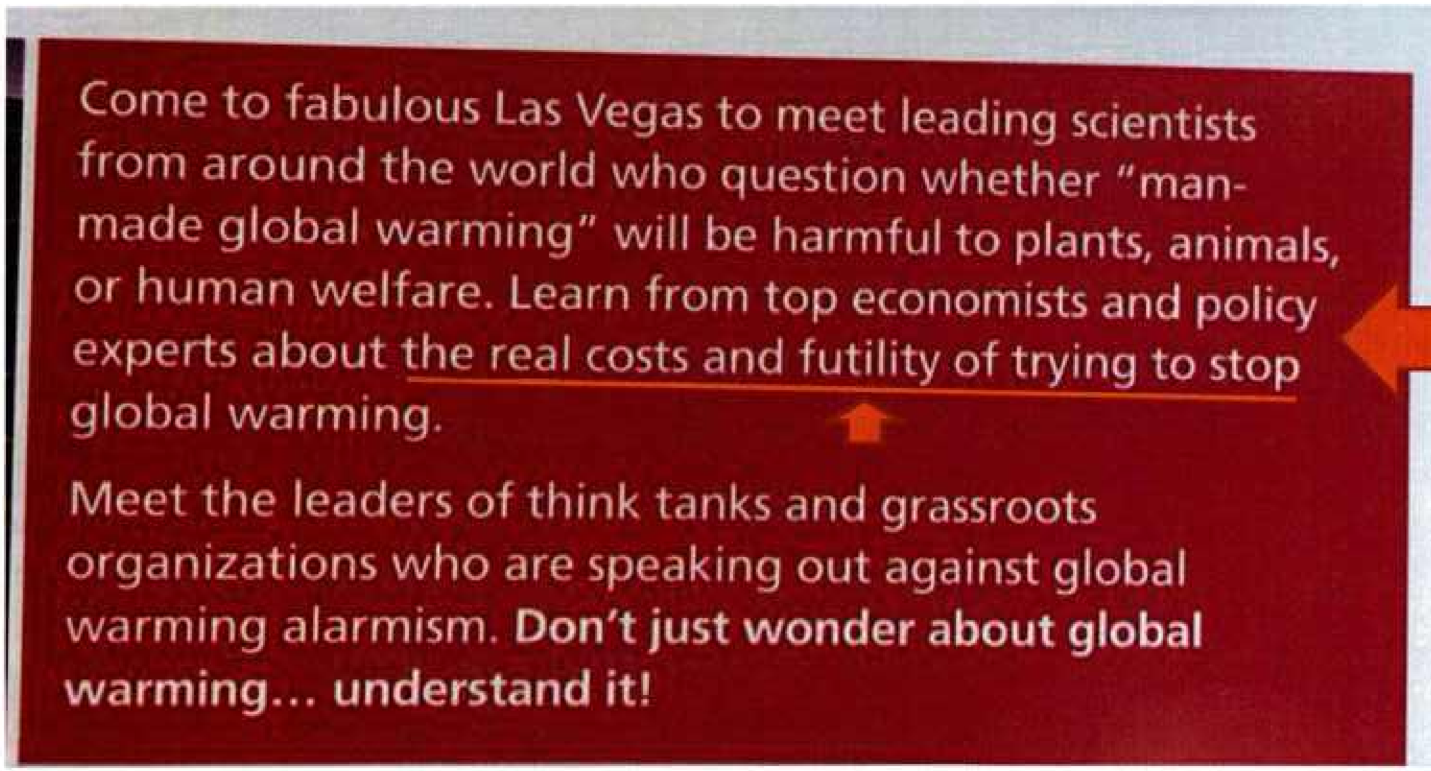
As a scientist who has studied the issue of climate change closely for some four decades, I can say with confidence that the author of this book has accurately
nd extent of
le global

nterpiece of
dren and our
his acceptance
n as the
began to heal."

lorful, often
with global
Climate
hed in
ate Change
NIPCC series
ne.

/D and

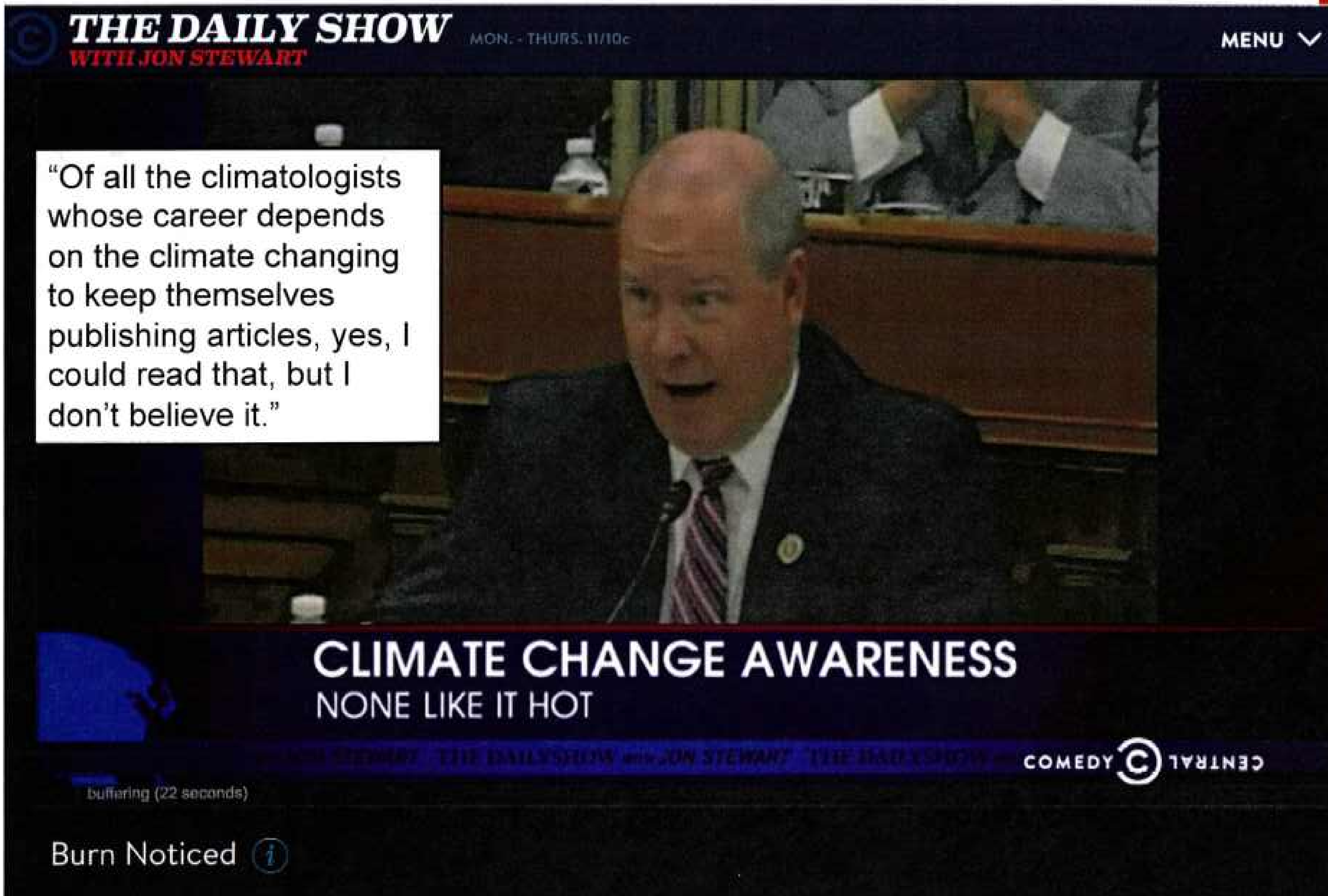
I have also received a handful of brochures, and invitations to a conference.



Come to fabulous Las Vegas to meet leading scientists from around the world who question whether "man-made global warming" will be harmful to plants, animals, or human welfare. Learn from top economists and policy experts about the real costs and futility of trying to stop global warming.

Meet the leaders of think tanks and grassroots organizations who are speaking out against global warming alarmism. **Don't just wonder about global warming… understand it!**


THE DAILY SHOW
WITH JON STEWART
MON. - THURS. 11/10c
SEARCH THE DAILY SHOW
LL EPISODES
VIDEOS
EXTENDED INTERVIEWS
GUESTS
NEWS TEAM
PODCAST
TICKETS
COMEDY CENTRAL
Burn Noticed


THE DAILY SHOW
WITH JON STEWART
MON. - THURS. 11/10c
MENU
“Of all the climatologists whose career depends on the climate changing to keep themselves publishing articles, yes, I could read that, but I don’t believe it.”
CLIMATE CHANGE AWARENESS
NONE LIKE IT HOT
COMEDY CENTRAL
buffering (22 seconds)
Burn Noticed


THE DAILY SHOW
WITH JON STEWART
MON. - THURS. 11/10c
MENU
Jon Stewart: "I cannot stress this enough. This is the House Committee on Science, Space, and Technology."
StumbleUpon
How long will it take for the sea level to rise 2'? I mean, think about it... if your ice cube melts in your glass, it doesn't overflow. It's displacement. This is some of the things they're talking about, that mathematically and scientifically don't make sense.
REP. STEVE STOCKMAN
(R) TEXAS
Play
07:44 / 10:25
Burn Noticed

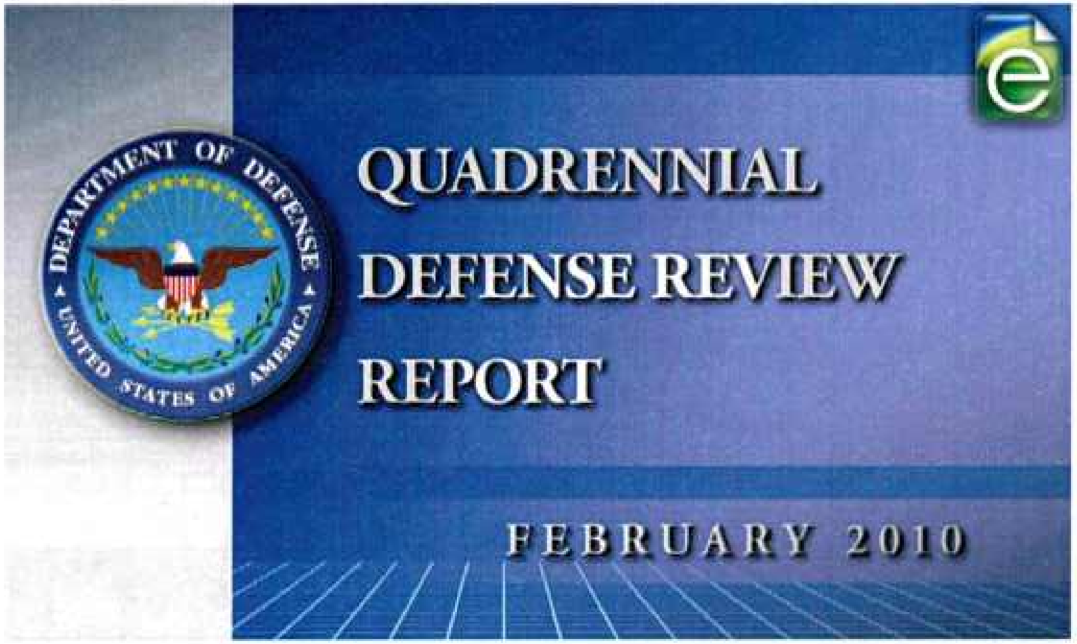
...a glass of ice water... doesn't overflow... *the ice is already in the water!*

The problem is ...there is other ice... that's on the *land!*

From a blog post on brookings.edu:
Security risks: The tenuous link between climate change and national security

"When I listen to the 'know-nothing' crowd and their front men in Congress who actively ignore ever-stronger scientific evidence about the pace of climate change, I want to quit my day job and organize civic action to close them down. The celebration of anti-knowledge, the denial of science, the treatment of advanced education as a mark of ignominy rather than the building block of American innovation and citizenship—these are as grave a threat to America's future as any I can identify."

— *Bruce Jones, Brookings Institution*



"Climate Change, Energy Security, and Economic Stability are inextricably linked. Climate change will contribute to food and water scarcity, will increase the spread of disease, and may spur or exacerbate mass migration." *—p. 107–108, QDR*

Climate change is a threat multiplier.

Living With Extreme Weather

A Workshop to Integrate Understanding and Improve Societal Response



http://extremeweather.ou.edu/index.html


EMERGENCY MANAGEMENT

SUBSCRIBE NEWS MAGAZINE EVENTS WEBINARS BLOGS PAPERS JOBS DEGREES ADVERTISI

Disaster Preparedness & Recovery

Central Texas Floods Expose Gaps in High-Tech Warning Systems

Technological and psychological issues with emergency alerts will be examined.

Eric Dexheimer and Sean Collins Walsh, Austin American-Statesman | June 8, 2015

"But he sort of brushed it off. He's been through several floods, and he wasn't worried. In fact, he later went to bed." .... By the time the Taylors were plucked from their home by a rescue boat, at about 3 a.m. May 24, the water had risen to their second floor.

http://www.emergencymgmt.com/disaster/Central-Texas-Floods-Expose-Gaps-High-Tech-Warning-Systems.html


PARKING LOT.
IF THIS
30
SEVERE T-STORM
WARNING

NEW METHODS ARE OVERDUE

Climatologists are bringing in analysis techniques from epidemiology. It’s not the average of climate that affects things, its the *extremes*.

<u>Extremes</u>: it’s how climate change is affecting us

There are <u>clear signals</u> in the extremes

- hot and cold temperature extremes on daily and seasonal timescales
- extreme precipitation events over most of the mid-latitudes will become more intense and more frequent

What new approaches are needed in the social, psychological, and decision-making realms?


THE SHIP CAN'T BE SINKING
MY END JUST ROSE 200 FEET

Emerging challenges for NDM: The case of security screening

Thomas C. Ormerod

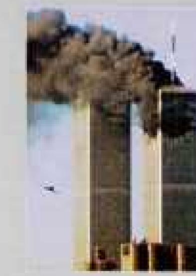
University of Sussex

NDM 2015, MITRE Corporation

Funding: UK MoD, EU, US DHS

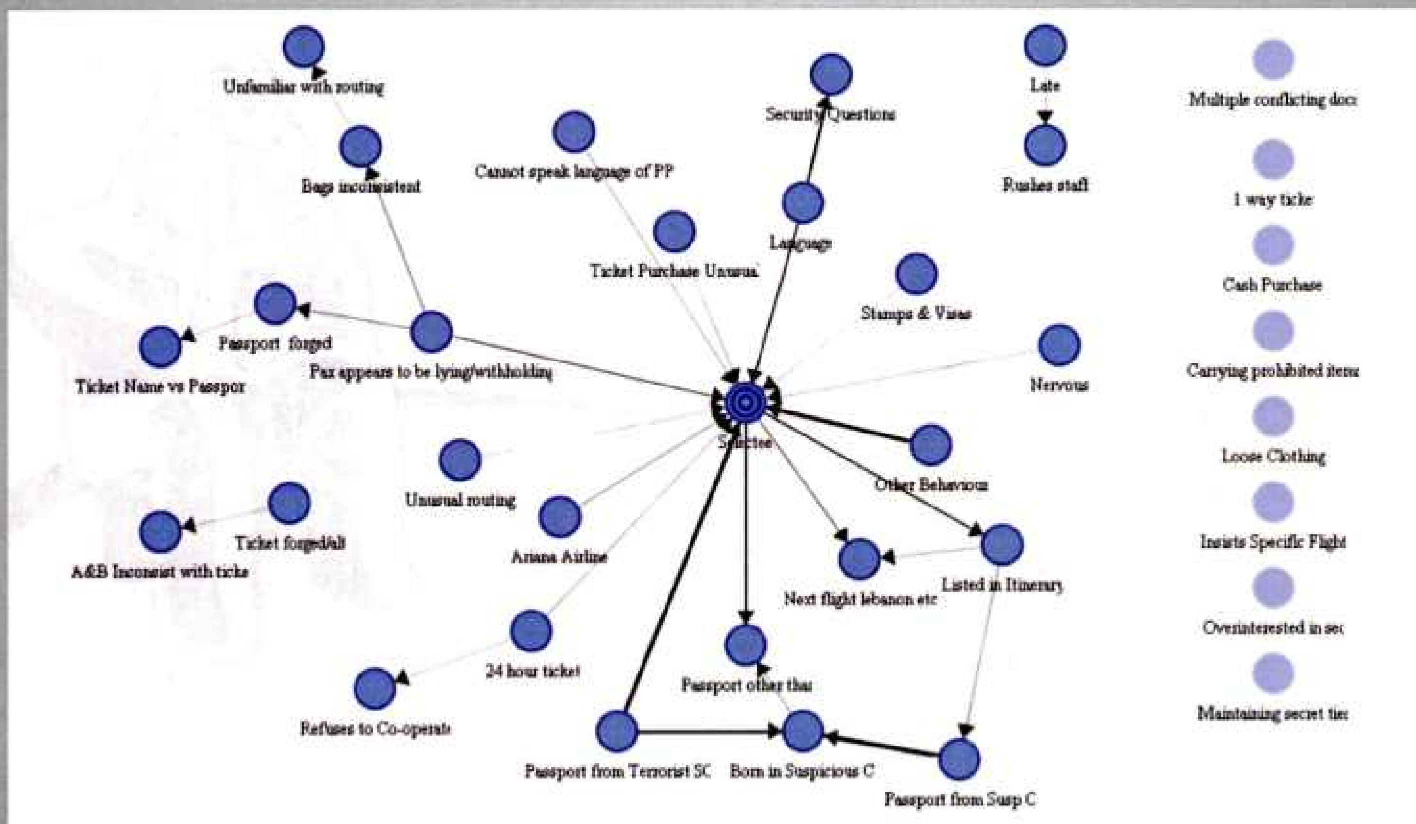
Current approaches to security screening

- Post 9-11, behavioral methods for threat detection have been based on Suspicious Signs
  - AOSSP Chapter 11
  - SPOT



- There are concerns but not enough evidence
  - Incidents
  - Forensic psychology evidence

Aviation Screening Study (2008)



- RL10 47-48 No check

Security agent stops an IC5 (late teens- 20
hasn't heard any of the conversat
Why is he here? He is stu
university. Who s
?
The pax s
around,
A
much longer will

pped him as he thought he could be
n interest in nuclear counter-proliferation.
ctually PRC but the same colour passport.

The pax is subsequently spotted in the departure lounge, he runs off on recognizing the security agent
After a search he is found with a different passport, different clothes and a package of money

Specifying a New Approach

- The aviation study tells us:
  - Avoid 'over-resolving' suspicions
  - Make veracity testing an explicit goal

- Forensic literature tells us:
  - Allow the interviewer to listen and watch
  - Unpredictability – interfere with the 'lie script'
  - Make the sender 'work' – Cognitive load

Controlled Cognitive Engagement (CCE) ™

- **Controlled**
  - Screener controls the conversation
  - Incremental phased questioning
  - Clear exit points
- **Cognitive**
  - Screener decision-making skills
  - Asymmetric cognitive loading
  - Unpredictable
- **Engagement**
  - Enhanced customer service
  - Reducing stereotype biases
  - Timeline to observe behaviour change.

→"Confidence to fly in three minutes"

Evaluation: Detection testing

- $500k field trial
  - Major EU hub airports
  - Two major international carriers
- Aim
  - To compare detection rates for CCE and suspicious signs method
  - To test method under pressure
- Method
  - CCE training
    - 10 accredited CCE trainers & 80 accredited CCE screeners
  - Double-blind randomised-control trial
  - 200 participants per method
  - Diverse participant sample (non-stereotyped)
  - Participant-generated deceptions
  - Incentivised performance

Results: Detection Rate

	January 2012	June 2012
CCE	63.4%	74.1%
Current method	2.7%	0%

- ◆ 4,000,000⁺ passengers CCE-screened to date
- ◆ Paedophile ring disrupted!

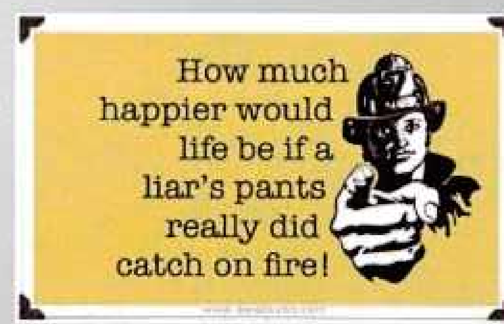
Challenge 1: Shifting NDM from ‘expert’ to ‘everyone’

1. Everyday expertise
2. Vulnerability & Ageing
3. Cultural differences

Challenge 2: Balancing human & technological inputs

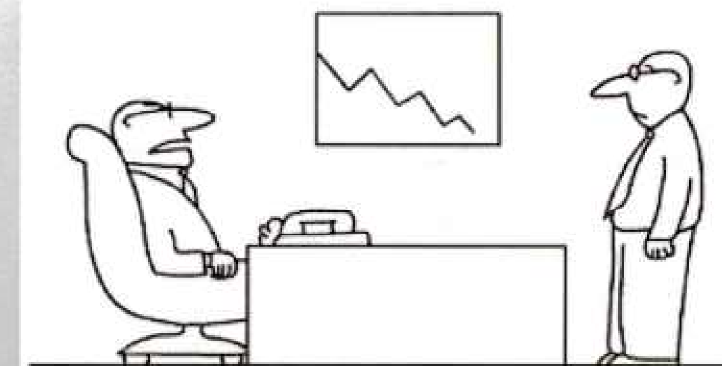
- On the one hand:
  - Computer-based face & document processing is better than human performance
  - Potential for remote covert detection
  - Potential to de-bias procedures

- On the other hand:
  - Machines give false confidence, create new vulnerabilities, and de-skill → 'irony of automation'
  - Current technologies cannot detect behavior *change*
  - Deceit is embedded in truth – targets must be verbally challenged



Challenge 3: Towards the next NDM theoretical framework

- RPDM captures major aspects of expert decision-making
- In security domains, the problems faced involve rare events with little or no feedback
- Yet there is skilled decision-making out there!
- Most decision-making is shared not owned.



"It would appear, Hopkins, that your gut feel was only indigestion"

Releasing the Adaptive Power of Human Systems

RE-

RE-

re-vision

re-framing

re-planning

re-engagement, ...

RE- re-vision
re-framing
re-planning
re-engagement, ...

already busy juggling multiple demands,
focused somewhere,
modifying plans in progress,
coping with complexity

RE- re-vision

re-framing

re-planning

re-engagement, ...

already busy juggling multiple demands,

focused somewhere,

modifying plans in progress,

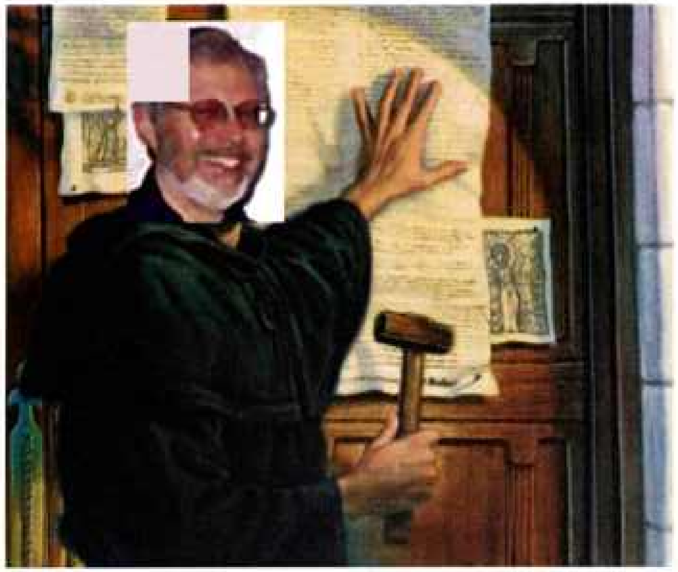
coping with complexity

and then things change ...

Naturalistic / Decision Making

30 years of effort = deja-vu all over again

Sisyphus-ian trap



DM [& the NDM reformation] **stuck**

- oversimplification fallacies
- decompensating in the face of change
- fragmented, mis-synchronized
- stuck in stale approaches

Feltovich and Spiro's **Oversimplifications**

Factors that are really:	Are treated as if they are:
Dynamic	Static
Continuous	Discrete
Interdependent	Separable
Simultaneous	Sequential
Heterogeneous	Homogeneous
Multiple representations/ functions	Single representations/functions
Systemic	Isolated
Nonlinear	Linear
Conditional	Universal
(regularly) Irrregular	Routinizable (regular)
Deep	Superficial

Naturalistic at scale

3 Drivers of Change & Innovation

1. **Connectivity**
2. **Sensors**
3. **Automation/Autonomy**

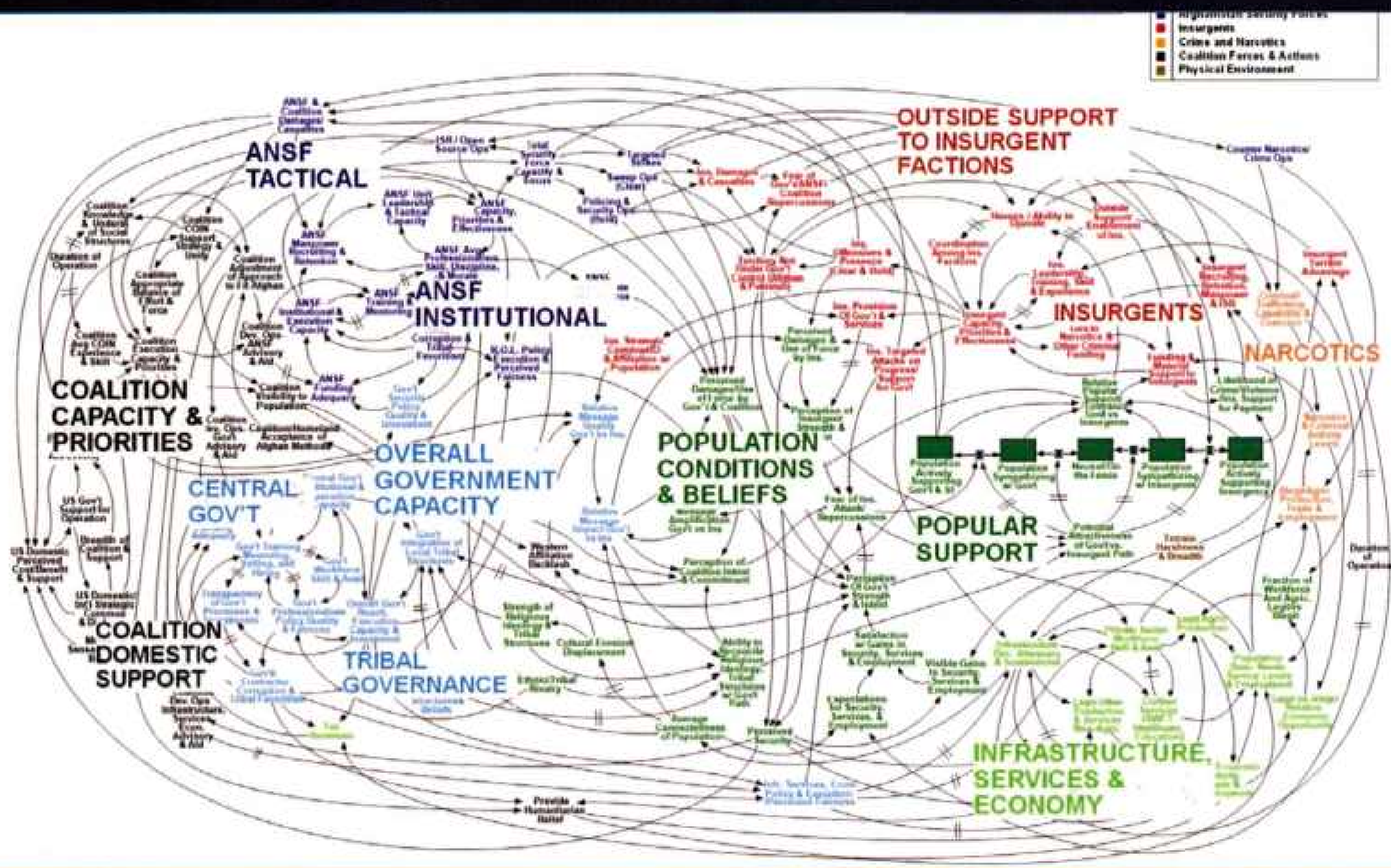
oops, ***4*** Drivers

1. Connectivity
2. Sensors
3. Automation/Autonomy *(scale effect)*

#4. People

human goals and expertise

law of unintended consequences
surprising cascades of effects
sudden failures
linear simplifications don't work anymore



Extensive and Hidden Interdependencies

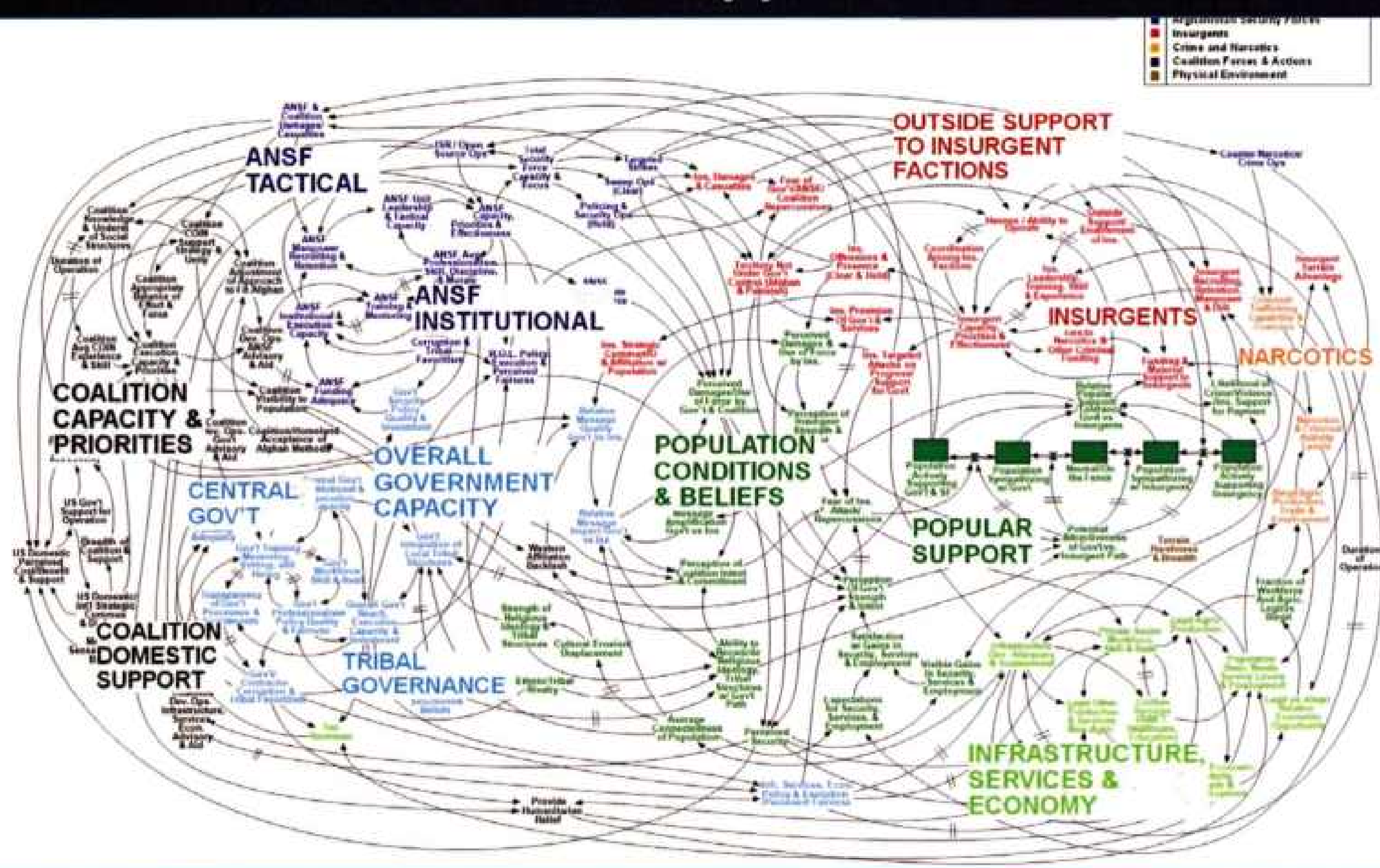
Surprising reverberations in tangled layered network

Unintended Consequences Offset Gains

"Even if the world were perfect, it wouldn't be"

Yogi Berra

Adaptive universe: You live in it,
pressures and technologies tempt you in,
you (and they) are trapped by its rules
it doesn't work the way you think it does



Tangled Layered Networks

Tangled layered networks

Boundaries

Surprise

Brittle

Decompensation

Graceful Extensibility

Fundamental Trade-offs

Adaptive Cycles

Polycentric Governance

Net Adaptive Value

Range of Future Adaptive Action

Cascades

Capacity for Maneuver

Risk of Saturation

Acute-Chronic Trade-off

Optimality - Brittleness Trade-off

Initiative

Reciprocity

Adaptive Cycles/Histories

empirical, general patterns

from collecting, sharing, analyzing stories of resilience & brittleness in action

cycles of [co-]adaptive reverberations across networks and scales

Precarious Present

assess where and how system is brittle/resilient in the face of surprise

risk of 3 patterns of adaptive breakdown

project how change produces unintended consequences

Resilient Future?

how to engineer in graceful extensibility and sustained adaptability given trade-offs and change?

Adaptive Cycles/Histories

- stories of resilience and brittleness in action
- extract general patterns

Critical Care

- Being Bumpable - Intensive Care Units (ICUs)
- Patient Boarding in Emergency Rooms

Strong Silent Automation

- Asymmetric lift incidents and accidents in aviation

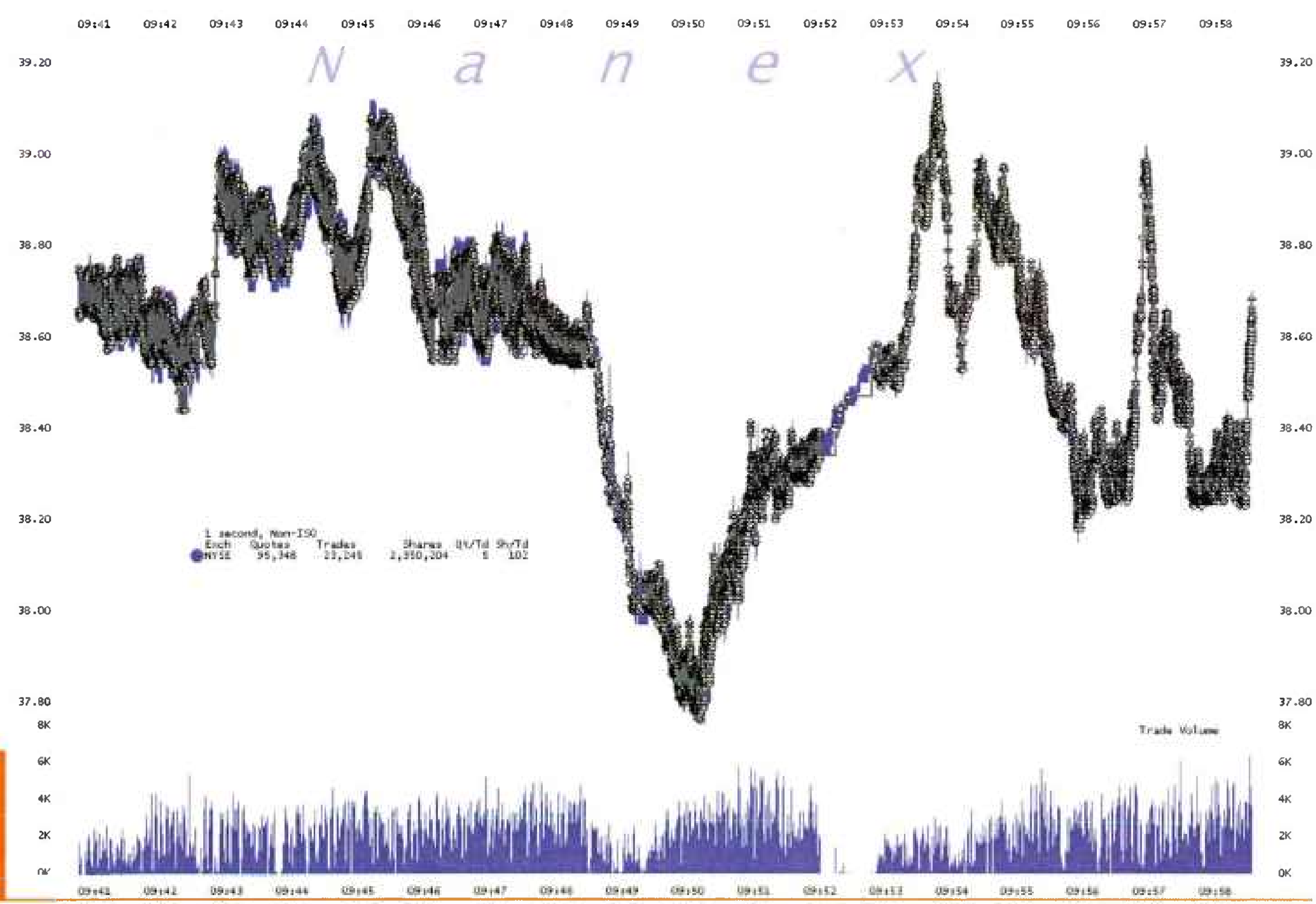
Business-critical digital services

- High-frequency trading;
  regular "flash crashes"

...

adaptive capacities

co-adaptive cycles reverberating across networks and scales



The Ohio State University

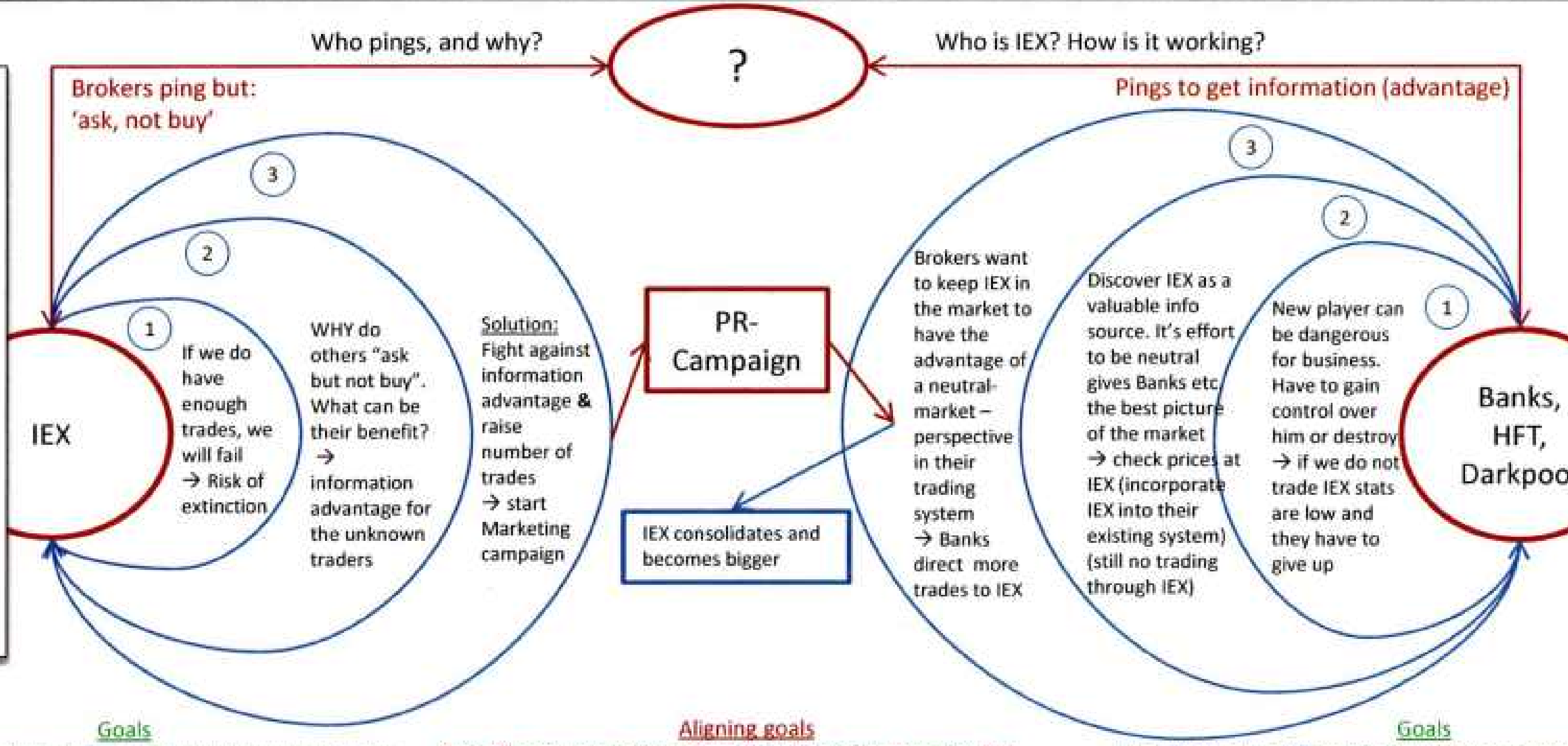
Law of Stretched Systems

With change every system is continuously stretched to operate at its new capacity.

People as problem holders exploit 'improvements' to better achieve goals by pushing the system out to operate near the edge of its new capacity boundaries. The process of adapting to exploit the improvement results in a new intensity, complexity, and tempo of activity.

Adaptive Cycle after an episode out of 'Flash Boys': Trigger - Launch of IEX

Background
IEX launches the first and only stock exchange that has included an anti-gaming model.
IEX is highly vigilant and anticipatory how the market reacts/adapts on its launch. Monitoring itself. Tries to find new predatory methods that were unknown and adapt.



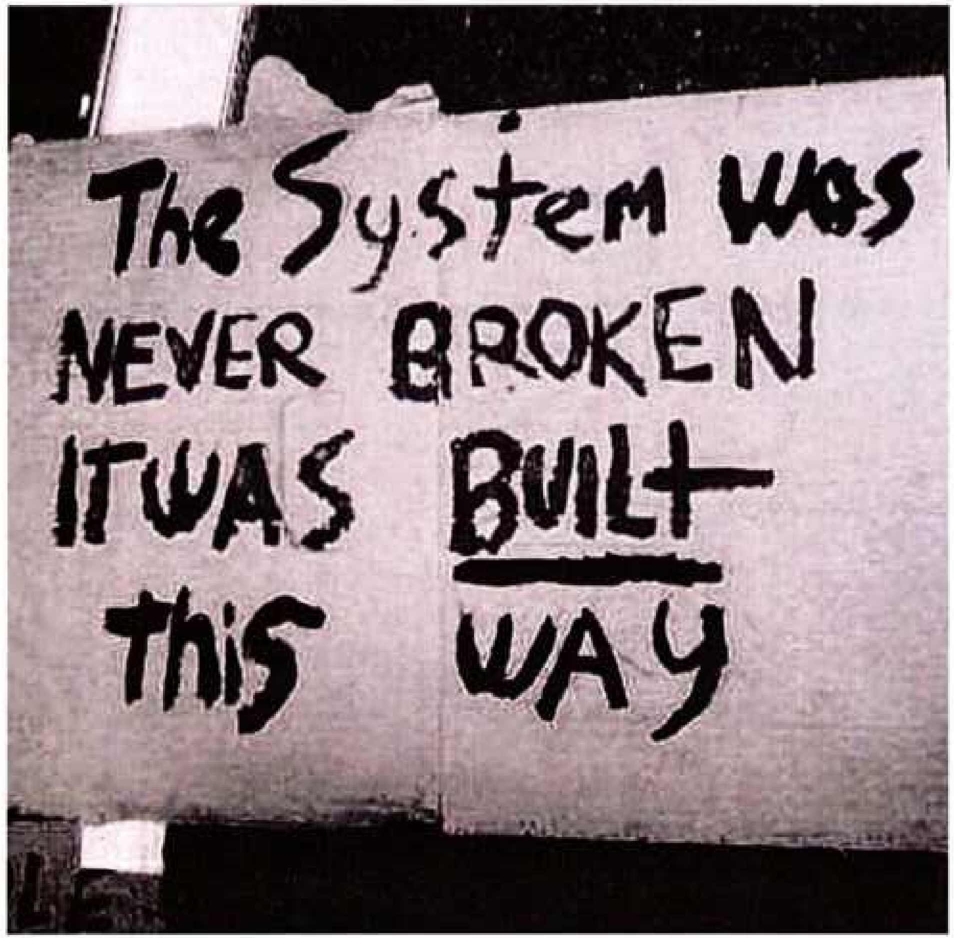
Background
Banks, HFT, and DP are UAB in this TLN. So are their employees and customers. They struggle for fitness, aiming for the next big deal, bonus or promotion. Keeps them going to adapt and create new possibilities to succeed and be better than others.

Goals
Survive on the market → being the best (shown by high trading stats to attract traders)
Establish an all new 'fair-trade-market' that is proven anti-gaming → constant struggle for fitness and tackling surprise

Aligning goals
Both IEX and Banks need trades to happen in order to survive. The Banks see that the anti-gaming market of IEX reduces the risk of losing money on a trade where they did not have the information advantage.

Goals
Make money by being better and smarter as competitors in the TLN of the market. Anticipate behaviour of other traders and then front running the market is the most popular technique. Having an information advantage is essential.

Charted by Jan Weber for ISE 5870 SS 2015



A system does what it is designed to do,
except that is not what the designer intended.

Adaptive Universe:

What is needed to be sufficiently adaptive & resilient as challenges change?

Graceful Extensibility

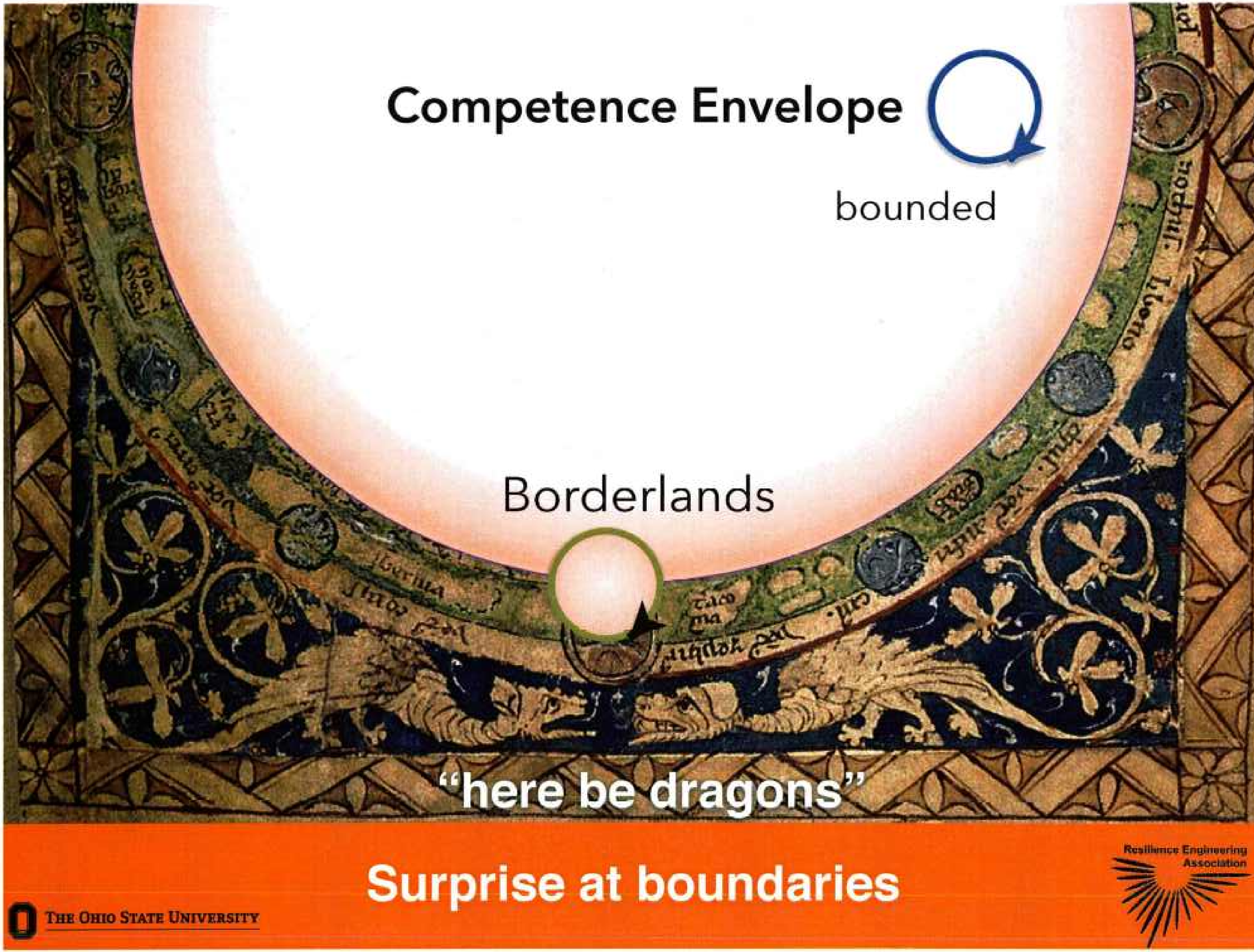
- graceful extensibility is a positive capability to stretch near and beyond boundaries when surprise occurs;
- graceful extensibility trades off with robust optimality
- change lead to shortfalls, responsible people make up the shortfall

Graceful Extensibility in Action

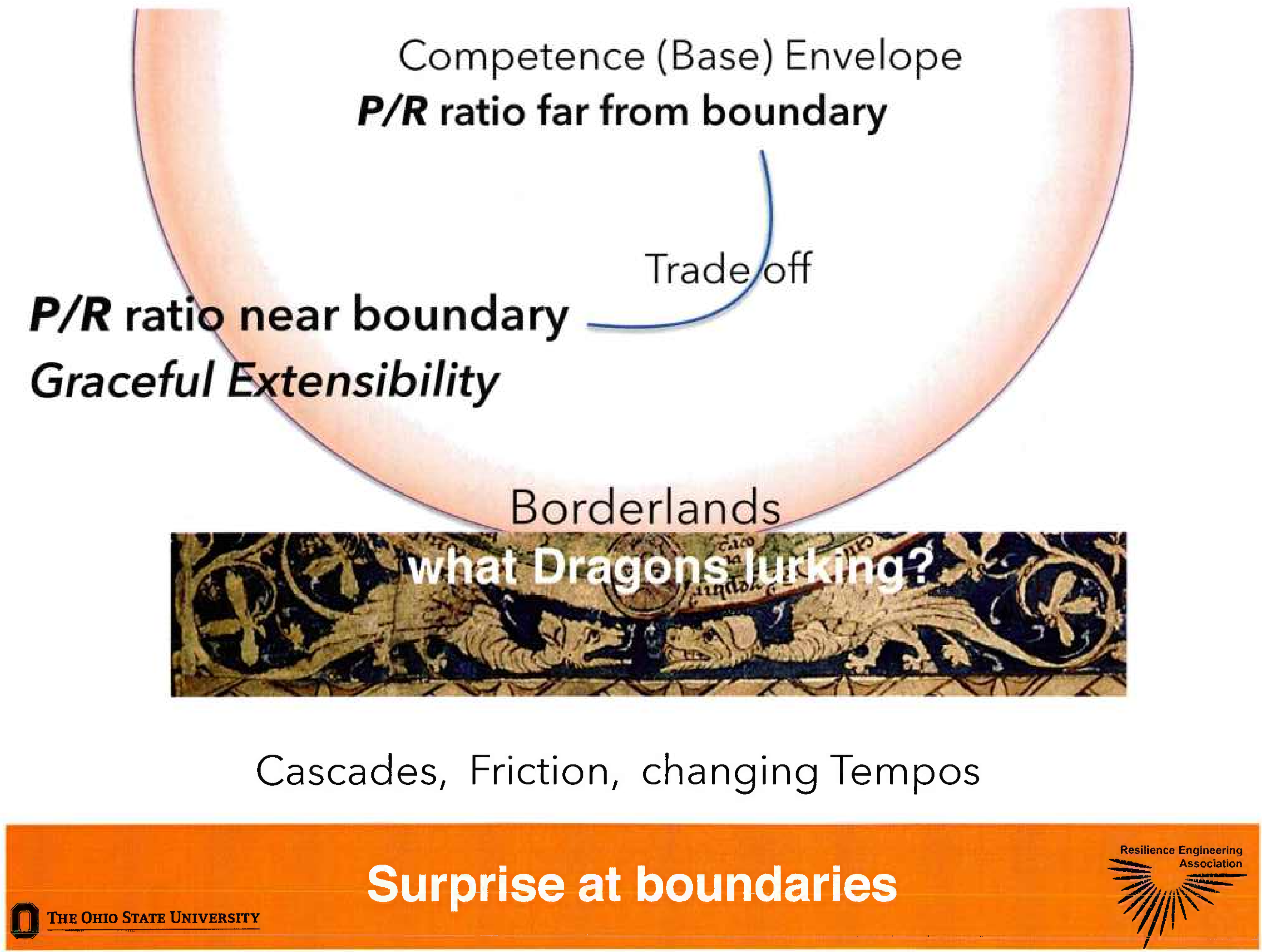
Surprise: there be dragons

The Ohio State University

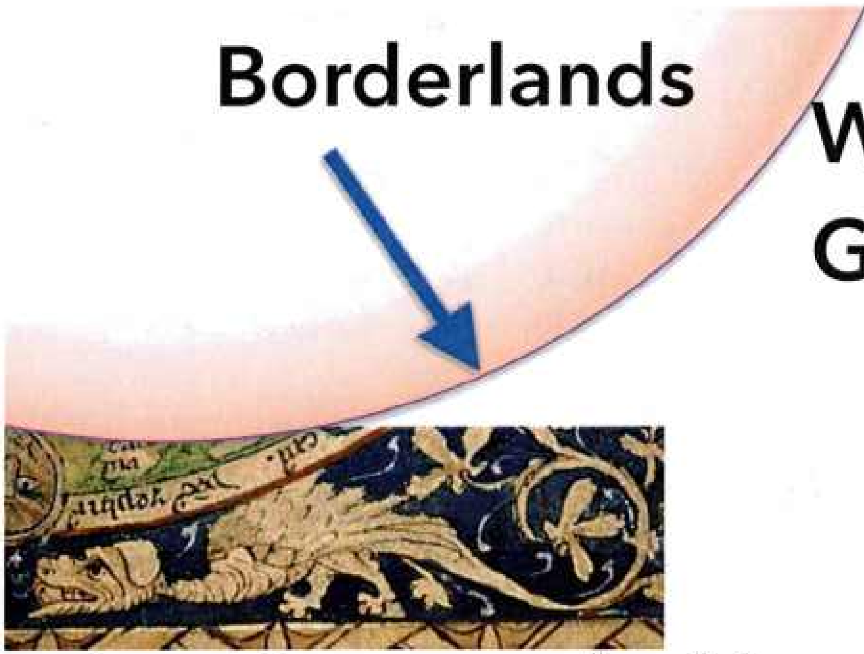
Resilience Engineering Association


Competence Envelope
bounded
Borderlands
"here be dragons"
Surprise at boundaries
THE OHIO STATE UNIVERSITY
Resilience Engineering Association

Surprise at boundaries



Competence (Base) Envelope
P/R ratio far from boundary
Trade off
P/R ratio near boundary
Graceful Extensibility
Borderlands
what Dragons lurking?
Cascades, Friction, changing Tempos
Surprise at boundaries
The Ohio State University
Resilience Engineering Association



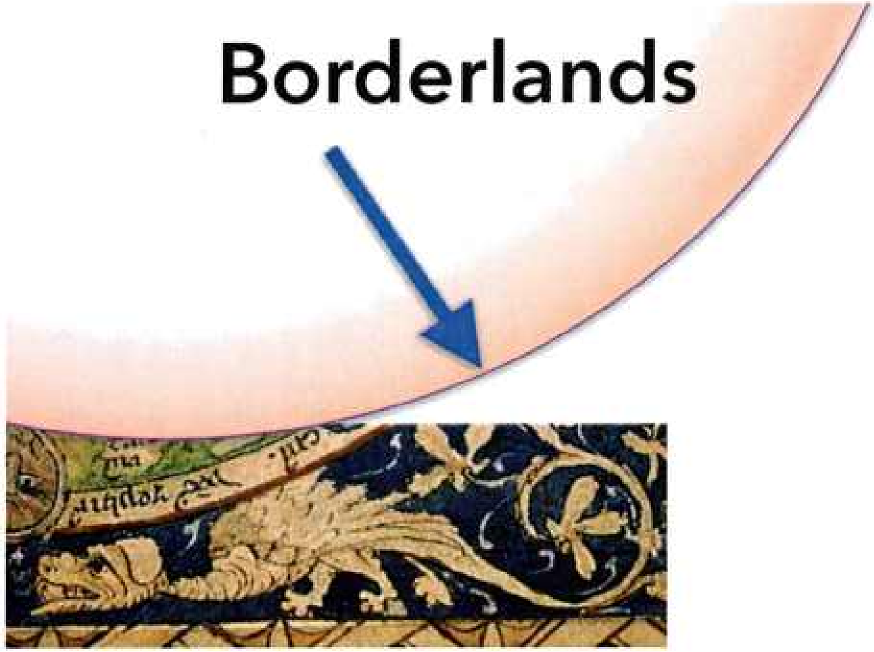
What adaptations produce Graceful Extensibility?

(descriptive) **Brittleness**

what Dragons lurking near, at and beyond boundary?

Anticipating crunches ahead

generating/sustaining a readiness to respond



Decompensation /Anticipation

Working at cross purposes / Synchronizing over units

Stuck in Stale / Proactive Learning

Competence Envelope

Far from

Near to **Borderlands**

Pursuit of optimality under FBC pressure

Adapt to sustain graceful extensibility

Net adaptive value

Balance of two P/R ratios

Competence Envelope

Borderlands

challenges to
handle Surprises

steepness of transition

Smooth effortless

Steep Gradient

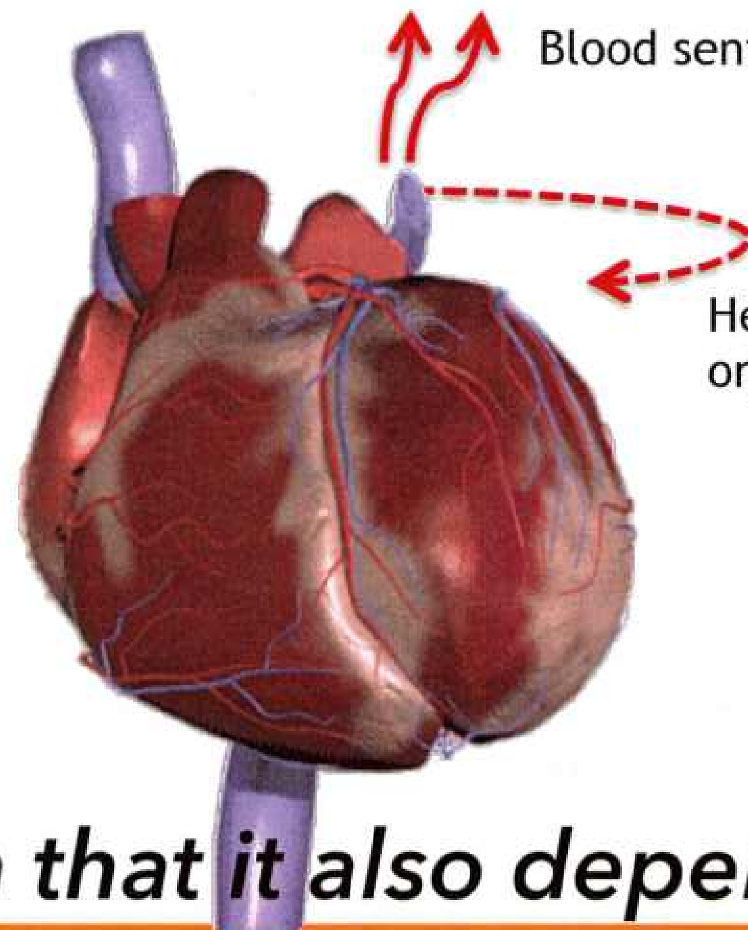
THE OHIO STATE UNIVERSITY

Resilience Engineering Association

with faults / load, goal conflict / cascades

keep work **low** to reduce stress

keep work **high** to supply O2



provides function that it also depends on

Graceful Extensibility

build the capacity to handle **crunches** ahead

- **anticipatory signals shared,**
- **modify readiness to respond,**
- **proactive learning**

all UABs have bounds and capacity saturates

key parameter:

Capacity for Maneuver (*CfM*)

risk of saturation of *CfM*

Theorem 1: Adaptive capacity is finite (or Boundaries are universal).
UABs have a limited range of adaptive behavior (boundary on adaptive capacity as variation, disruptions, and change occur. All UABs can run out of adaptive capacity as the demands faced grow and change in difficulty.

Capacity for Maneuver (*CfM*) = the range of adaptive behavior, how much as been used & how much remains as unit responds to changing demands.

Theorem 2: Events will produce demands that challenge boundaries on the adaptive capacity of any UAB (or Surprise occurs continuously).
All UABs)risk *saturation* - running out of *CfM* as upcoming events present increasing demands.

Theorem 3: Adaptive capacities are regulated to manage the risk of saturating *CfM* (or Risk of saturation is monitored and regulated).
UABs adjust/regulate their adaptive capacities relative to the risk of saturating *CfM*.

Theorem 4: No UAB can have sufficient ability to regulate *CfM* to manage the risk of saturation alone (or coordination over multiple UABs in a network are necessary).

Coordination and alignment across multiple UABs in a network is needed to extend the range of adaptive behavior to match changing and increasing demands.

Theorem 5: Some UABs monitor and regulate the *CfM* of other UABs in response to changes in the risk of saturation (or 2nd order Regulation of *CfM)*.

Some UABs monitor the risk of saturating *CfM* in another UAB by monitoring the increasing effort to stay in-control. When they recognize that the risk of of saturating the *CfM* is too high, they respond to expand the *CfM of the* UAB at risk.

Theorem 6: Adaptive capacity is the *potential* for adjusting patterns of action to handle future situations, events, opportunities and disruptions (or potential for future adaptive action).

Adaptive capacity exists before change and disrupting events call upon that capacity. Regulation of *CfM* provides capacity to respond to future challenges and opportunities.

Theorem 7: Performance of a UAB as it approaches saturation is different from the performance of that UAB when it operates far from saturation (or pursuit of optimality alone undermines extensibility).

Managing the risk of saturation of *CfM* (reducing brittleness to acceptable relative risk levels), at a minimum, requires two forms of interdependent adaptive capacity –'base' and 'extended' and *Net adaptive value* assesses both.

Theorem 8: All UABs are local.

All UABs are embedded in an environment and in a neighborhood of relationships across a portion of a network of UABs.

Theorem 9: There are bounds on the perspective any UAB, but these limits are overcome by shifts and contrasts over multiple perspectives (or perspective contrast overcomes bounds).

The view from any point of observation simultaneously reveals and obscures properties of the environment (there is no best perspective).

Theorem 10: Reflective systems risk mis-calibration (or mis-calibration is the norm).

Since risk of mis-calibration is omnipresent, effort must be invested to reduce risk of mis-calibration. In other words, since there is a bound on how well models of capability match actual capability, effort must be invested to improve the match.

Manage complexity? *tilting toward florescence*

- tangible experience with ***surprise***
- unease - our systems are ***precarious***
- human talent – ***initiative***: decentralized, coordinated
- invest resources to toward ***resilience as graceful extensibility*** – not only FBC pressure
- build ***reciprocity*** across roles, and units, and levels.

How do you manage complexity & tilt toward florescence as connectivity, sensing, automation capabilities grow?

positive story?	negative story?
graceful extensibility	brittleness
initiative / reciprocity	shortfalls & mis-calibration
adjust capacity to adapt	trapped in trade-offs

the basic rules to reach resilience are emerging to:

leverage the adaptive power of human systems

RE-


Re-Framing Human Systems
Co-
Adaptive
Complexity in Natural, Social and Engineered Systems

Welcome to MITRE

International Naturalistic Decision Making 2015

General Information

Entry:

This meeting will be held in MITRE building M1 in the Hayes Auditorium (7525 Colshire Drive, McLean, VA 22102) Conference entry will be through the South Lobby.

MITRE will provide meeting badges for all participants on-site in McLean. **Please have a valid photo ID (government ID, driver's license or Passport for foreign nationals) available at check in**.

Refreshments:

Continental breakfast, lunch and an afternoon snack will be provided; please advise of any dietary restrictions.

Logistics:

Information regarding hotel, directions, parking and the Metro shuttle schedule can be located on slides 9-13.

International Naturalistic Decision Making 2015

Points of Contact

Attendance, Logistics and General Assistance:

Anita Maginniss 703.328.2633 (cell)
amaginniss@mitre.org

Meeting planning and content questions:

Gary L. Klein 703.983.5291 (office)
gklein@mitre.org

Agenda – Tuesday, June 9

8:30	**Continental Breakfast & Registration**
9:30	**Welcoming Remarks**
	Dr. Mark Maybury, VP & CTO The MITRE Corporation
	Dr. Robert Hoffman, Institute for Human and Machine Cognition
	Dr. Gary L. Klein, Conference Chair, The MITRE Corporation
10:00	**Opening Keynote Session**
	Scott Tousley, Deputy Director of the Cyber Security Division, U.S. Dept. Homeland Security Science & Technology
11:00	**Session 1: Aviation**
	Reducing the Information Scavenger Hunt to Improve Air Traffic Management Decision-Making, **Catherine Bolczak and Amanda M. Staley**
	Applications of a Prioritization Methodology for Complex Operational Problems and Solutions, **Zheng Tao**
12:00	**Lunch**
13:00	**Session 2: Teams**
	The Challenges of Asynchronous Communication for Distributed Teamwork: Task Performance and Media Effects, **Kathleen Mosier, and Ute Fischer**
	Externalizing Planning Constraints for More Effective Joint Human -Automation Planning, **Ronald Scott, Emilie Roth, Beth Depass and Jeffrey Wampler**
	The Influence of Operating Room Handoffs on Teamwork, Stress, and Work: a 360 Degree Evaluation of Team Shared Situation Awareness, **Cristan E. Anderson, MD and Lygia Stewart, MD**
	Naturalistic Decision Making, Risk Management, and Leadership, **MAJ. Bridgette Bell and COL. James Ness**
14:45	**Break**
15:00	**Session 3: Methodology**
	(Very) Rapid Decision Making: Framing or Filtering?, **Chris Baber , Xiuli Chen and Andrew Howes**
	New Practices for Simulation Metamodeling in a Modern Computing experimentation Environment, **Scott L. Rosen, Jim Ramsey, Christine Harvey and Samar K Guharay**
	Crowdsourcing Mental Models using DESIM (Descriptive to Executable Simulation Modeling), **Mark S. Pfaff, Jill L. Drury, and Gary L. Klein**
	Tools for Facilitating Critical Decision Method during Tacit Knowledge Elicitation, **Danny Shu Ming Koh, Hock Guan Tee, Boon Kee Soh and Angela Li Sin Tan**
16:15	**Break**
16:30	**Session 4: Sports**
	Professional Judgement and Decision Making in Sport Coaching: To Jump Or Not To Jump, **Andy Abraham and Dave Collins**
	Assessing Hitting Skill in Baseball using Simulated and Representative Tasks, **Patrick K. Belling, Jason Sada and Paul Ward**
	From Lab to Cage: Turning the Occlusion Research Method into a Sports Training Program, **Peter Fadde**
18:00	**Reception**
19:00	**Adjourn**

Agenda – Wednesday, June 10

8:30	**Continental Breakfast**
9:15	**Keynote Speaker**
	Dr. Alvin Roth, Craig and Susan McCaw professor of economics at Stanford University and the Gund professor of economics and business administration emeritus at Harvard University: *Market design as a process of adjustments: the evolution of kidney exchange, 2000-2015*
10:15	**Break**
10:30	**Session 5: Culture**
	Good Stranger Diagnostic Tool: Measuring Capacities and Limitations to Inform Training, **Helen Altman Klein and Joseph Borders**
	Revealing and Assessing Cognitive Processes Underlying Cultural Acuity Through Domain-Inspired Exercises, **Matthieu Branlat, Julio C. Mateo, Michael J. Mccloskey and LisaRe Brooks Babin**
11:30	**Session 6: Training**
	An Empirical Evaluation of the ShadowBox Training Method, **Gary A. Klein, Joseph Borders, Corinne Wright and Emily Newsome**
	Aiding Police in the Detection of Imminent Terrorist Attacks: Testing Different Approaches to Improving Situation Assessment Skills, **Ben Morrison and Natalie Morrison**
12:30	**Lunch**

13:30	**Session 7: Safety Panel**
	Improving Safety: What are Leverage Points from a Macrocognition Perspective? **Emilie M. Roth, Laura G. Militello, Rhona Flin, and Austin F. Mount-Campbell**
14:45	**Break**
15:00	**Session 8: Intelligence Analysis & Cyber**
	Distributed Sensemaking: A Case Study of Military Analysis, **Simon Attfield, Bob Fields, Ashley Wheat, Rob Hutton, Jim Nixon, Andrew Leggatt and Hannah Blackford**
	Applying the Principles of Magic and the Concepts of Macrocognition to Counter-Deception in Cyber Operations, **Simon Henderson, Robert Hoffman, Larry Bunch and Jeff Bradshaw**
	The SAnTA Recommender System and Naturalistic Deci-sion Making **Frank Linton, Sarah Beebe, Mark Brown, Casey Falk, and Mark Zimmermann**
	Agility in Allocating Decision-Making Rights for Cyberspace Operations, **Steven W. Stone**
	How Military Intelligence Personnel Collaborate on a Sense-Making Exercise, **Chris Baber , Gareth Conway, Simon Attfield, Chris Rooney, Neesha Kodagoda and Rick Walker**
17:00	**Adjourn**
18:30	**Optional: Social Recognition Prime Dinner at Seasons 52 Restaurant**

Social Recognition Prime Dinner

Seasons 52 FRESH | GRILL®

Located in the Tysons Corner Center Mall, 1st Level

Seasons 52 has the fresh appeal of a farmer's market, serving ingredients at their peak of ripeness, prepared with rustic cooking techniques like brick-oven roasting and grilling over an open fire of oak and mesquite. And with an award-winning wine list carefully crafted by our Master Sommelier, and fresh cocktails infused with seasonal fruits and herbs, there's always something new to discover.

Please see Anita Maginniss or Gary Klein if you wish to register for this event.

Agenda - Thursday, June 11

8:30	Continental Breakfast
9:00	Session 9: New Domains, New Theories
	Identifying Critical Cues in Mental Health Assessment using Naturalistic Decision-Making Techniques, **Ben Morrison, Julia Morton and Natalie Morrison**
	Intuitive Potential and Predicting Entrepreneurship – a Study on a New Method of Measuring Intuition, **Matylda Gerber**
	Evidence-Based Decision Making in Civilian Agencies: An Analysis of Three Cases, **Suzanne L. Geigle**
	Expertise and decision making in real estate appraisal: results from a naturalistic study, **Olga Preveden**
11:00	Break
11:15	Session 10: Sports 2
	Exploring Cue Use in Rugby League Playmakers to Inform Training Initiatives, **David Johnston and Ben Morrison**
	What is the Story Behind the Story? Two Case Studies of Decision-making under Stress, **Anne-Claire Macquet and Héloïse Lacouchie**
	Look Who's Talking - In-game Communications Analysis as an Indicator of Recognition Primed Decision Making in Elite Australian Rules Football Umpires, **Timothy J. Neville and Paul M. Salmon**
12:45	Lunch & Keynote Speaker
	CAPT Joseph Cohn, Deputy Director, Human Performance Training and BioSystems (HPT&B) Directorate: *Understanding, Representing & Enhancing Intuitive Decision Making: From Individuals to Societies -- Progress, Challenges, and Opportunities for Representing Behavior*
14:00	Break
14:15	Session 11: Methodology 2
	Expertise Management: Challenges for adopting Naturalistic Decision Making as a knowledge management paradigm, **Brian M. Moon, Holly C. Baxter, and Gary Klein**
	Advancing ACTA: Developing Socio-Cognitive Competence/Insight, **Julie Gore, Adrian Banks and Almuth McDowall**
15:15	Invited Speaker
	Marvin Cohen, Principal Investigator at Perceptronics Solutions
16:15	Break
16:30	Poster "Lightning Round" Introductions
17:00	Poster Session with Cash Bar
18:30	Conference Banquet
	Guest Presentations on the History of NDM

Agenda – Friday, June 12

8:30	Continental Breakfast
9:00	Panel Session: Emerging Challenges
	Mr. John Willison, Director, Command, Power, & Integration, U.S. Army RDECOM CERDEC
	Professor Tom Ormerod, Head of Psychology at the University of Sussex
	Dr. Chris Baber, Chair of Pervasive and Ubiquitous Computing, School of Electronic, Electrical and Systems Engineering, University of Birmingham, United Kingdom
10:15	Break
10:30	Panel Session: Emerging Challenges 2
	Simon Henderson, Centre for Cyber Security & Information Systems
	COL Matthew Hepburn, DARPA Program Manager -- dynamic threats of emerging diseases
	Daphne Ladue, Research Scientist with Oklahoma University's Center for Analysis and Prediction of Storms
	Jeff Bradshaw, Sr. Research Scientist, Institute for Human and Machine Cognition, *Cyber-Physical Threats in the Food Industry: Toward Real-Time Anticipation, Detection, Response, and Recovery*
12:00	Lunch & Invited Speaker
	Dr. David Woods, Integrated Systems Engineering at the Ohio State University -- *Releasing the Adaptive Power of Human Systems*
13:15	Break
13:30	1/2 - Day Tutorials
	Gary A. Klein -- Cognitive Task Analysis
	Robert Hoffman -- Concept Mapping
	Gary L. Klein -- Decision Spaces and Option Awareness
17:30	Conference Adjourns

Hyatt Regency Tysons Corner Center

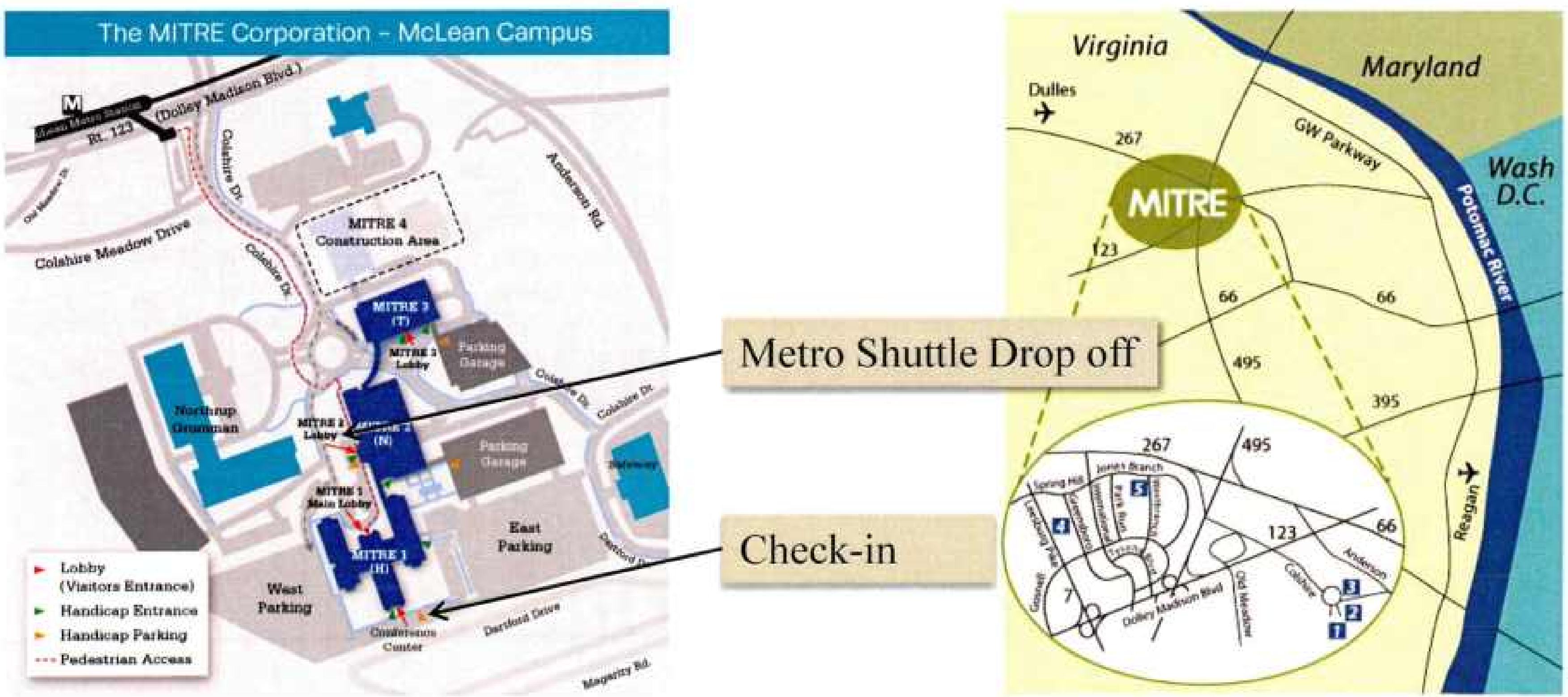
•300 guestrooms, including 18 suites, Regency Club rooms, and Presidential Suite
•Access to the 300+ stores and restaurants of Tysons Corner Center regional shopping mall via covered walkway
•Closest hotel to the all-new Tysons Corner stop of the Silver Line Metrorail (steps away)
•15 miles from Washington Dulles International Airport, Ronald Reagan National Airport and downtown Washington, D.C.
•Our refreshing indoor heated pool and 24-hour Stay Fit gym allow you to maintain your exercise routine
•Discover a world of dining options at Tysons Corner Center steps away from the hotel

Hyatt Regency Tysons Corner Center
7901 Tysons One Place
Tysons Corner, Virginia, USA, 22102
www.tysonscornercenter.regency.hyatt.com
Tel: +1 703 893 1234

International Naturalistic Decision Making 2015

Location

Check-in for the conference will begin at 7:30 am in the south lobby located at the MITRE 1 Building (H). The MITRE 1 Building is located on the McLean MITRE campus at 7525 Colshire Drive, McLean, VA.



Directions to MITRE-1

- **From Reagan National Airport**
- Take George Washington Parkway North approx. 6 miles to Route 123 South, McLean exit.
- Exit onto Route 123, (also called Dolley Madison Blvd.) follow south for approx. four miles to traffic light at Colshire Drive (on left).
- Turn left onto Colshire Drive and continue through circle on Colshire.
- Parking is available in the East and West parking lots.
- Enter the South Conference Center lobby.

\- or -

- Take I-66 West. Take Exit 67 to Route 267 (Dulles Toll Road).
- Take Exit 19A, following signs to Route 123 South (also called Dolley Madison Blvd) follow south for approx. 1/4 mile to traffic light at Colshire Drive (on left).
- Turn left onto Colshire Drive and continue through circle on Colshire.
- Parking is available in the East and West parking lots.
- Enter the South Conference Center lobby.

From Dulles Airport

•Take the Dulles Airport Access and Toll Road to Exit 19. Take Exit19A, following signs to Route 123 South, Tysons Corner.
•At the light at the end of the ramp, turn right onto Route 123 (also called Dolley Madison Blvd.).
•At next traffic light turn left onto Colshire Drive and continue through circle on Colshire.
•Parking is available in the East and West parking lots.
•Enter the South Conference Center lobby.

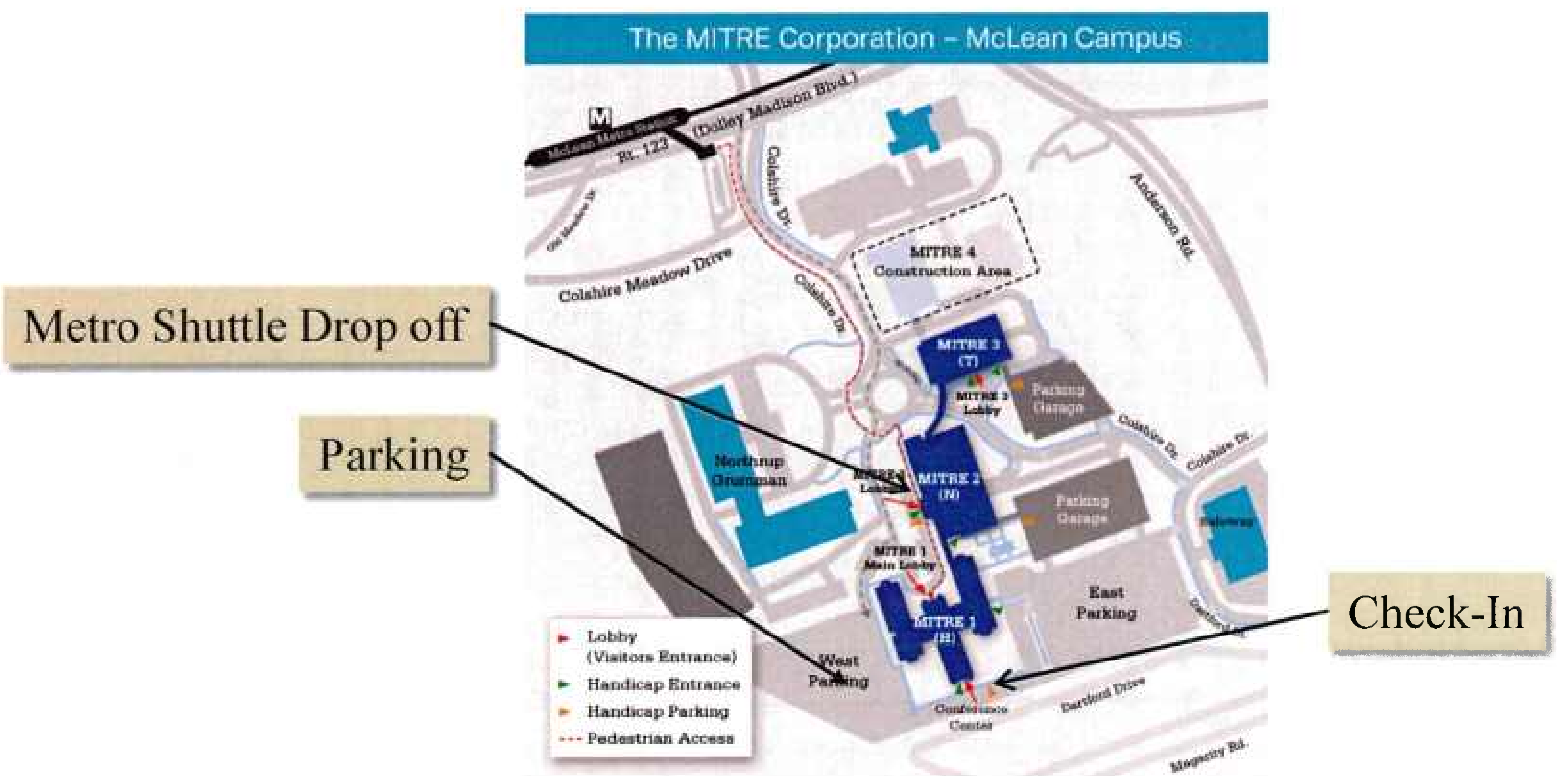
From Route 495 (Beltway)

•Take Exit 46B (McLean, Route 123). Take Route 123 North, (also called Dolley Madison Blvd) go to the second traffic light at Colshire Drive.
•Turn right on Colshire Drive and continue through circle on Colshire.
•Parking is available in the East and West parking lots.
•Enter the South Conference Center lobby.

Parking

Park in the West parking lot and walk to the Conference Center entrance on the South side of the MITRE-1 building for access to the International Naturalistic Decision Making 2015.

If using the Metro Shuttle, walk to the Main Lobby (North side) of MITRE-1 for admittance



Silver Line Metro Shuttle

For the convenience of MITRE's employees and visitors with business-related activities (such as meetings, interviews, and conferences), we provide a shuttle service from the kiss-and-ride area of the McLean station on Metro's Silver Line throughout the workday.

The 11-passenger, ADA-compliant van will travel from the McLean station to three different MITRE locations:

All times listed are departure times. The first shuttle leaves the McLean station kiss-and-ride at 6:45 a.m. weekdays. The last pickup from MITRE 2 (at 5:55 p.m.) will drop off passengers at the McLean station before the shuttle goes off-duty.

Printed schedule cards are available at all building reception desks and on board the shuttle.

Departure times from each location.				
Loop A		**Loop B**		
McLean Metro	**Rappahannock**	**McLean Metro**	**Polk**	**MITRE 2**
6:45a	7:00a	7:15a	7:20a	7:25a
7:30a	7:45a	8:00a	8:05a	8:10a
8:15a	8:30a	8:45a	8:50a	8:55a
9:00a	9:15a	9:30a	9:35a	9:40a
9:45a	10:00a	10:15a	10:20a	10:25a
10:30a	10:45a	11:00a	11:05a	11:10a
11:15a	11:30a	11:45a	11:50a	11:55a
12:00p	12:15p	12:30p	12:35p	12:40p
12:45p	1:00p	1:15p	1:20p	1:25p
1:30p	1:45p	2:00p	2:05p	2:10p
2:15p	2:30p	2:45p	2:50p	2:55p
3:00p	3:15p	3:30p	3:35p	3:40p
3:45p	4:00p	4:15p	4:20p	4:25p
4:30p	4:45p	5:00p	5:05p	5:10p
5:15p	5:30p	5:45p	5:50p	5:55p

NOTE: The last pickup from MITRE 2 (at 5:55p) will drop off passengers at the McLean Metro station before the shuttle goes off-duty.